Who will benefit from immunotherapy?
Can we develop a pre-treatment predictive model?

PHC Catalyst

ImmuunPRO Hackathon

Transforming the fight against cancer using Big Data & AI

An experiment using data from the randomized OAK and POPLAR trials

'Imagine the future is now: just do it'

June 2020
Accelerating the transition towards PHC by creating a receptive environment. This study was commissioned by the PHC Catalyst Alliance with financial support of Roche Netherlands B.V.

Personalised Healthcare Catalyst

This report is developed by the members of the PHC Catalyst Alliance and has been made possible by a financial contribution of Roche Nederland B.V.

Project team

PHC Catalyst Alliance

Ewoudt van de Garde
Michel van den Heuvel
Daan Hurkmans
Paul Iske, Chairman

Roche NL

Rob Bohte
Cor Sibum
Judith van Schaik, PHC Business Development Lead

Accenture NL

Laury van Bedaf
Gaston Besanson
Thijs van den Broek
Ana Faraon
Kirsten Spahr van der Hoek, Innovation Lead

For further information please contact ***info@phc-catalyst.nl***

Foreword

This report is the outcome of a project initiated by the Personalised Healthcare Catalyst.
The PHC Catalyst Alliance started in 2018 with the mission to accelerate the transition towards personalised healthcare (PHC), as rapid progress in biomedical, data science and AI create new insights into health and disease that enable us to transform patient's lives by *delivering care tailored to the individual*, thereby helping to prevent, diagnose, and treat patients more effectively and quickly.

Though the progress in the various scientific areas is impressive, we do not fully enjoy the benefits yet, primarily due to the barriers in the receiving environment, in particular the way we currently have organized our healthcare system. In other words, science is running faster than its framework.

The PHC Catalyst Alliance is a 'coalition of the willing'; it is a multi-disciplinary group of professionals, each experts and innovators in their fields, who are united in their vision and ambition to create a healthcare system without barriers for PHC. This will allow healthcare to evolve from reactive 'one-size-fits-all' disease care towards proactive personalised healthcare, and will allow consumers to use personalised health information to improve their health as they observe the impact of their lifestyle decisions.

In 2020, the Personalised Healthcare Catalyst Foundation was established, a legal entity with the same purpose as the alliance. We look for opportunities to highlight and accelerate progress, and initiate and support projects to do so. We apply the methods of 'Combinatoric Innovation', which means that we do not reinvent the wheel, but focus on what is already out there and connect people, organisations, knowledge and data to demonstrate the value of PHC, shift mindsets, and break down implementation barriers.

This report is dedicated to one of these projects. Actually, this is how the PHC Catalyst started off, with an idea to perform a datahackathon to develop a pre-treatment predictive model for immunotherapy in patients with lungcancer, based on the combination and analysis of data from various sources. ImmuunPRO Hackathon focuses on transforming the fight against cancer using Big Data & AI. Immunotherapy is a major scientific breakthrough in cancer treatment with the potential to change cancer in a chronic condition with a good quality of life. Currently only a small percentage of patients with metastatic cancer has a durable response. The key question is: Who will benefit from immunotherapy? In other words: why does one person have a durable response and the other not? This type of knowledge will eventually help us to further improve treatment.

Interestingly, it turned out that finding, accessing and combining data sources, were challenges by themselves. This proves the point that reshaping the system with its shared, but also individual and conflicting ambitions and interests, is key. I therefore compliment and thank all involved in this project, not only for the scientific work they have done, but also for the stamina to convince people and organisations that it is worthwhile to share and participate.

This report is one in a series that we are developing and hope that the results will inspire people to further develop and implement the knowledge that has been acquired during the process. But we also hope that more people and more organisations will work together on changing the society and the healthcare system in particular, to make personalised healthcare affordable and accessible for everyone.

United we stand, departed we fall!

Kind regards, Paul Iske, Chairman PHC Catalyst

1 Executive summary

Who will benefit from immunotherapy?[1]

Cancer immunotherapy is a significant improvement in the treatment of cancer. Novel patterns of response and progression to immunotherapy have been reported, that are not observed with conventional cytotoxic or targeted anticancer treatments. The major breakthrough with immunotherapy is its potential to achieve durable responses in a subset of patients with advanced cancer that can be maintained several years even after stopping the treatment. However, a substantial proportion of patients does not respond to immunotherapy.

Immunotherapy drugs called immune checkpoint inhibitors (ICI) work by blocking checkpoint proteins that are made by some type of immune system cells, such as T cells, and some cancer cells. These checkpoints help keep immune responses from being too strong and sometimes keep T cells from killing cancer cells. When these checkpoints are blocked, T cells can kill cancer cells better. Examples of checkpoint proteins found on T cells or cancer cells include PD-1/PD-L1.

When doctors and patients consider ICI treatment, the survival outcome is uncertain. Theoretically, patients with high PD-1/PD-L1 expression levels are more likely to respond to PD-1/PD-L1 inhibitors that block the activity of PD-1/PD-L1 immune checkpoint proteins. However, in clinical practice patient outcomes vary considerably. This raises the question about a potential role of PD-1/PD-L1 expression as predictive biomarker for the selection of patients to treat with ICIs. A pre-treatment predictive model for survival outcome may be useful for the provision of more personalized expectations of survival outcomes, which may inform treatment choices and contribute to a higher quality of life.

Experimental design

This experiment aimed to develop and validate a pre-treatment predictive model for survival outcomes in advanced non-small cell lung cancer (NSCLC) patients treated with ICIs.

By combining different information about the specific traits of a patient (Big Data), and with the help of artificial intelligence (AI) and the expertise of medical experts, it is possible to develop such a pre-treatment predictive model. The key question is whether the pre-treatment predictive model is good enough to be used in clinical practice.

The training and testing data we used for developing the model consisted of clinical and pathological data from 1512 patients that were enrolled in the phase II POPLAR and phase III OAK trials. These patients with previously treated NSCLC were randomly assigned to receive atezolizumab, a PD-L1 inhibitor, or docetaxel. The primary outcome in our experiment was defined as overall survival (OS) after 2 years (yes or no).

1. Borcoman et al. Annals of Oncology 30: 385-396, 2019.

Predictors and outcomes

The following five variables were most predictive for OS after 2 years (in descending order of predictive value): C-reactive protein (CRP), number of metastatic sites, time since diagnosis, neutrophils, PD-L1 expression. For PFS after 3 months additional predictive variables were found, the most important of those was CD19.

The OS model after 2 years has an AUC of 0.81. The area under the curve (AUC) measures the performance of the predictive model. In general, for diagnostic tests an AUC of 0.5 suggests no discrimination, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.[1]

The OS after 2 years seems to be influenced by the use of antibiotics. In depth machine learning analysis revealed a difference between atezolizumab versus docataxel. Further analysis will be done to see if the number of antibiotic courses is important in this respect.

Conclusions

A pre-treatment predictive model was developed which discriminates patients with distinct differences in survival outcomes following atezolizumab (PD-L1 inhibitor) treatment for advanced NSCLC.

For patients considering the initiation PD-L1 inhibitor treatment of advanced NSCLC the pre-treatment predictive model presented for OS after 2 years may be useful for the provision of more personalized expectations of survival outcomes, which may inform treatment choices and contribute to a higher quality of life.

Clinical relevance

This experiment showed that pre-treatment CRP is an important prognostic variable for OS after 2 years in advanced NSCLC patients treated with a PD-L1 inhibitor. Together with other routinely collected data: number of metastatic sites, time since diagnosis, neutrophils and PD-L1 a fair prediction can be made for the individual patient regarding OS after 2 years (yes or no). Taken into account the variables used in the models it is remarkable that PD-L1 expression is of limited value for predicting the treatment outcome. A general conclusion is that a simplification, the use of a single biomarker for prediction (e.g. PD-L1 expression), underestimates the complexity of diseases and is of limited value. Big data, especially relevant deep data, is needed to further improve our understanding of the underlying cause(s) of disease in the individual.

1. Mandrekr J. J Thorac Oncol. 5: 1315-1316, 2010.

2 Highlights of the experiment

2.1 Introduction

In daily clinical practice doctors and patients are challenged to make decisions about treatments. One question which often arises is, "will the drug work for me doctor"? The objective of this experiment is, to make a model that predicts the outcome of treatment with immunotherapy by using Machine Learning in patients with advanced Non-Small Cell Lung Cancer (NSCLC). The members of the team that took the challenge were Dutch Healthcare professionals (HCP) and data-researchers, all members of the PHC Catalyst Alliance. The vision of this alliance is to accelerate the use of personalized healthcare in daily clinical practice. The initial thoughts were to organize a data-hackathon to tackle the questions involved.

2.2 Data access (see also Appendix A: Data access journey)

Interestingly, finding, accessing and combining data sources, turned out to be by far the hardest challenge in this experiment: it took us nine months to get access to data. The next barriers were found:

- *GDPR concerns:* according to the GDPR guidelines data used for study purposes has to be completely had to be anonymized. Effective data anonymization is made up of two parts:[1]
  - It is irreversibel.
  - It is done in such a way that it is impossible (or extremely impractical) to identify the data subject.
- *Data mining:* is difficult because current practice is that patients give informed consent for a specific research question (testing hypothesis), therefore the data cannot be used for datamining (generating new hypothesis). Datamining involves exploring and analyzing large amounts of data to find patterns of correlation which may generate hypothesis on health/disease understanding;
- *Data pollution:* clinical trial data are of very high quality, therefore there is a genuine concern that combining these data with other data from other sources (e.g. hospital data) may reduce the reliability of the insights generated from data (garbage in, garbage out);
- *Data sources considered:*
  - Roche clinical trial data;
  - Flatiron real world data: at the time of the experiment Flatiron data was only accessible for researchers and clinicians that participate in the Flatiron hospitals and academic centers network;
  - Dutch hospital data: is only accessible for specific research questions and not yet for datamining. The process is very time-consuming, and (perceived) privacy concerns make the management of hospitals reluctant to participate in this type of innovative experiments;
  - Open Dutch data sources (e.g. BBMRI): are only accessible for researchers and the process is similar to what is described above for the hospitals, thus extremely difficult and time-consuming.

After two years parts of data sets from two studies, POPLAR and OAK, in which in total 1512 patient were enrolled (see appendix Immunotherapy for details). The PDMA department (Global Medical Affairs), together with a data scientist from gRED (Global R&D unit, Genetech) partly prepared and shared the filtered datasets (e.g. DNA sequencing data, exploratory biomarkers, pharmacokinetic/ pharmacodynamic data and radiology data were filtered out) with the Accenture data scientists. Of course, the data was anonymized according to both the GDPR and Roche guidelines.

1. https://gdpr.eu/?s=anonymizationation.

2.3 Contracts & IT

- *Contracts:* The internal Roche processes for selecting and contracting the vendor took four months. There is no internal process for data-mining (e.g. a data hackathon), and thus no legal contract template for this purpose. The alternative was a dataroom (a 'virtual' room containing secured date) where the data was analysed by dedicated data scientists from the vendor. We used an adjusted clinical trial legal template for this purpose. This covered confidentiality and any IP-issues.
- *IT:* Global IT and Accenture global IT discussed different solutions for the dataroom environment. Different cloud enviroments were heavily discussed (AWS, Google Cloud, Roche's own cloud), but the most pragmatic solution turned out to be storing the data on two local Roche laptops. The data could only be accessed by the data scientists from the vendor and the Roche Netherlands team members, so none of the other participants could access to the data. Their role was to guide the data analysis by giving feedback on the results.

3 Data-analysis

The steps followed during the data-analysis phase of the experiment:

3.1 Step I

Summaries of datasets of the two studies were reviewed and discussed with the data-analysis-group (PHC Alliance members) and the questions to address were fine-tuned, which led to a total of 12 research question. As a next step we selected three questions based on complexity and impact of the questions.

The three questions:

1. Can we build a robust model to predict if a patient survives after 2 years of treatment (OS)?
2. Can we build a robust model to predict if a patient has no cancer progression after 3 months of treatment (PFS)?
3. What is the effect of antibiotics on OS in patients treated with atezolizumab/ immunotherapy

Based on these questions a first selection of data was done (Table 2: Data sources). One has to understand that the outcomes in this approach are binary, i.e. OS at 2 years = yes or no; idem for PFS. At the moment immunotherapy is offered to the patient he or she has reached an advanced stage of cancer. Then the questions asked by the patient are usually about survival; "how long will I live" or "will I still be alive after X years". We focused on the later type of question for survival.
The timeframe of 2 years OS was chosen by the healthcare professionals in the data-team. Also, the PFS time window of 3 months was chosen because this is the average time it takes for the immune system to respond to the immunotherapy and start fighting cancer.

3.2 Step II

Exploratory data analysis and further preparation of the data were done to make an analytical base table (ABT). In this table all variables (of interest) are rearranged and linked to the outcome variable you want to predict (Appendix C: Modelling Variables).

During this process discussions took place with the HCPs for interpretation and selection of data. After this step the data is in a proper format to be 'processed' by an algorithm.
In parallel correlation analysis between variables and OS outcome were performed to check if more variables might be of use for modeling.

This resulted in 1407 observations and 441 features for PFS and 1379 observations and 437 features for OS (a feature is a processed variable in such a way that it can be used by the algorithm) (Table 5: Feature engineering based on variables from the data sources).

Hierarchically clustered correlation plots are available for each ABT component: basic patient characteristics, co-medication, lab values, medical history, and vital signs. These plots provide a qualitative overview of the correlations. (Figure 4. Hierarchical clustering of correlations between lab values and OS target and Appendix G: Correlation Analysis).

3.3 Step III

In the supervised machine learning (ML) modelling phase, different options were tested, some of them listed in Appendix B.

In the supervised machine learning (ML) modelling phase, different algorithms and modelling options were tried out to build the optimal predictive model for our scope. In the patient population there is a high number of features given the sample, therefore we use an approach based on features correlation with the target. A constraint, or penalty, is added when the model has too many features and thus features that are not informative for the model are removed.

We found this approach to perform best with the XGBoost algorithm in predicting OS at 2 years; the AUC = 0.81, sensitivity = 0.67 and specificity = 0.76, at a threshold of 0.35 (Table 7: Summary of ML models built for OS prediction).

For PFS at 3 months we use the same modelling approach that performs best with a GBM algorithm AUC = 0.64, sensitivity= 0.77 and specificity = 0.33 at a threshold of 0.5 (Table 8: Summary of ML models built for PFS prediction).

In total 27 features were used for the OS model (Table 6: Feature importance for the OS and PFS predictive models), the top 5 features model are:

1. CRP
2. Number of metastatic sites
3. time since diagnosis
4. neutrophils
5. PD-L1

For the PFS model 15 features were selected the top 5 being:

1. CRP
2. time since diagnosis
3. CD19
4. Neutrophils
5. Number of metastatic sites

Performance evaluation per treatment arm

In the modelling design, OAK & POPLAR patient data was used to train the OS model irrespective of the treatment arm. When we test the model for patients taken treatment into account, 8% of the ATEZ patients are falsely classified as non-survivor while for the docetaxel group this is 8.7%. The conclusion is that we assume there is no significant difference in the OS model performance between the two treatment arms.

Hopkins

Finally, we compared our modeling results with a study by Hopkins that was published during our data-analysis sessions.[1] Their research question was to predict OS based on OAK and POPLAR data used for training the model. Data sets from two additional studies (FIR & BIRCH) were used for external validation of the model. FIR & BIRCH were single arm atezoluzimab studies. Based on statistical analysis - multivariate Cox proportional hazard model - they defined five prognostic groups from low to high and estimated OS for these groups using Kaplan-Meier estimates. Although the study design between Hopkins *et al.* and our predictive model is different, it is possible to compare the important variables and performance of the two models.

1. Hopkins A.M. et al. Clin Cancer Res 26:3280-6, 2010.

The same variables were found to be important predictors in the Hopkins model and in our model. Hopkins paper used statistical analysis - univariate and multivariate Cox proportional hazard models to find the top predictors, while we used ML to generate this list seen in Table 6. Regarding performance of the Hopkins model compared to our model, which should be done with caution as explained above, Hopkins model reaches c-statistic 0.72 in the development cohort (OAK & POPLAR) and 0.76 in the test cohort (FIR & BIRCH). Our model reaches 0.81 when tested on the OAK & POPLAR population.

3.4 Step IV

The effect of antibiotics on OS.

A number of publications have discussed the possible negative outcome of check-point inhibitors due use of to antibiotics.

A recent paper by Chalabi, who performed a pooled ad-hoc analysis on the OAK and POPLAR study data and suggested that use of antibiotics (ABT) from 30 days before treatment until 30 days after start of treatment is associated with worse survival outcomes.[1]

To analyze the effect of ABT on overall survival a Kaplan-Meier was made (Figure 7: Kaplan-Meier survival plot). This clearly shows that patients on atezolizumab without the use of ATB have a better OS compared to antibiotic users ($p< 0.0001$). We are performing an analysis to see if the use of antibiotics during the treatment course with atezoluzimab also has a negative impact (so the use of antibiotics 30 days or more after randomization). Outcome is pending.

If the analysis shows that the use of antibiotics during the treatment with CIT will have a negative effect an outcome, this is an important finding.

3.5 Step V

Is our predictive pre-treatment OS model good enough?

Whether or not the OS prediction model with an AUC= 0.81 is sufficient for clinical use, is not clear. Depending on the patient's needs for certainty the model can support the choice of treatment.

However, we still need to validate model with real world data. Does the AUC remain 0.81 or will it decrease?

Some considerations regarding the validation:
PFS is more complex to model due to various factors including pseudo-progression and definitions used. The Hopkins paper also build a statistical model for PFS. Their model reached performance 0.60 on the development cohort and 0.61 on test cohort. Our model for PFS reached 0.64. While this is not robust enough to implement in the clinic yet, it shows the model has potential to be further refined with additional exploration & analysis of the data.

1. Chalabi M. et al. Ann Oncol 31:525-531, 2020.

3.6 Potential next steps

3.6.1 This model

- An additional analysis on the influence of the use of antibiotics on survival outcomes will be done (the results will be presented in the final version of this report). If the analysis shows that the use of antibiotics during the treatment with CIT will have a negative effect an outcome, this is an important finding.
- The model still needs to be validated. Ideally, we would like to use real-world data from Dutch hospital patients for external validation of the model.
  - The next step would be to collect retrospective anonymized data from patients with advanced NSCLC from one or more hospitals. Selection of the hospital(s) should be done by the team.
  - Another validation source could be The NVALT registry as this contains a lot of (unstructured) data of around 3.500 NSCLC patients treated with immunotherapy.
- Datamining of the current dataset enriched with deeper data to the datasets (e.g. -omics data, PK/PD data, imaging data) to improve or redefine the model.

3.6.2 In general

- Qualitative research to determine the value of predictive models in daily clinical practice: when is a model good enough according to the doctor and the patient?
- A desire to combine data from different databases: Pharmo/Palga/NVALT/IKNL (feasibility study is ongoing with 20 patients).
- A desire to develop model for immunotherapy in the earliest stage of lung cancer (does this data exist anywhere in the world?).

3.7 Conclusions

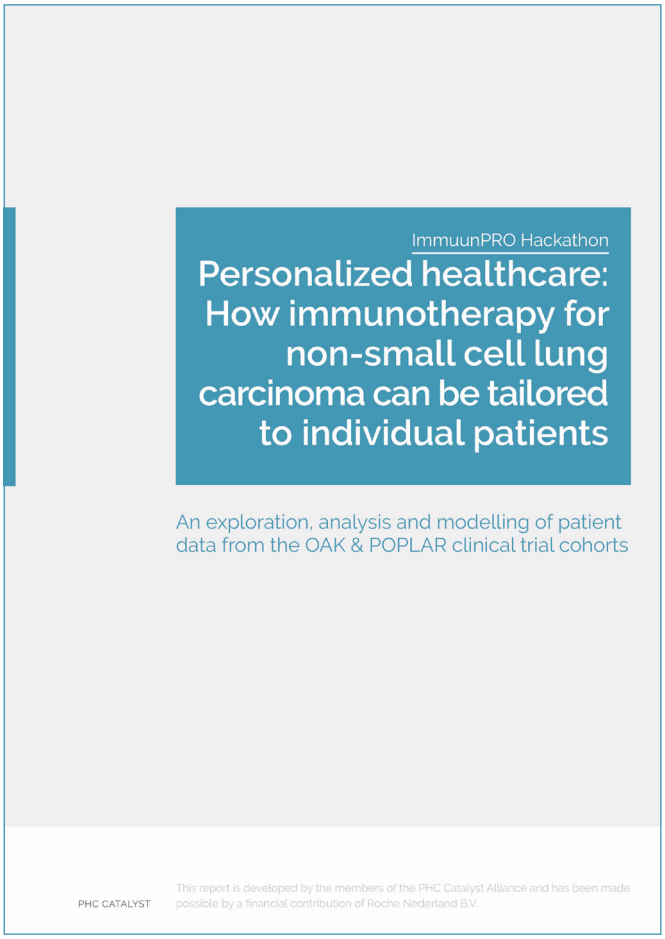
The collection of 'big data' to build prediction models via machine learning is a time-consuming exercise. Within Roche no clear processes are in place, so final decision making for providing data is difficult. Of note, our project has made this problem very clear and things are moving within the organization to have a clearer process in place. In Dutch hospitals the use of data for ML is even harder and hopefully the Dutch government will take an active role for guidance.

The model that was developed still needs improvement but shows that ML can be used as a decision support tool. Improvement of the model by adding more in-depth data is likely. However, this type of data is often not captured in daily clinical practice and needs investments for expensive diagnostic procedures.

The hurdles to overcome before the model can be tested in the hospitals with real world data are still in place. We will continue this journey, learning by doing is our way forward.

4 The Experiment in detail

Please follow this link to access the full ImmuunPRO data hackaton report:
[Personalized healthcare: How immunotherapy for non-small cell lung carcinoma can be tailored to individual patients]



5 Appendices

5.1 Immunotherapy a major breakthrough in cancer treatment

5.1.1 Disease complexity

Cancer is a complex multifactorial disease. Complex diseases are difficult to study and treat, because many specific factors that cause or contribute to the disease have not yet been identified. Every tumor is truly unique and each patient is different. This explains why not all patients respond to standard therapy, one size does not fit all. A more personalized approach is needed.

5.1.2 Immunotherapy

Novel patterns of responses[1]

Cancer immunotherapy is a significant advance in the treatment of cancer. Novel patterns of response and progression to immunotherapy have been reported, that are not observed with conventional cytotoxic or targeted anticancer treatments. The major breakthrough with immunotherapy is its potential to achieve durable responses in a subset of patients with advanced cancer that can be maintained several years even after stopping the treatment.

However, a substantial proportion of patients does not respond to immunotherapy. Furthermore, pseudoprogression occurs, supporting the concept of treating some patients beyond progression, hyperprogression occurs, where it is essential to interrupt the treatment and switch to another potentially active treatment, and finally some patients experience dissociated responses, with some lesions shrinking and others growing, where local treatment with surgery or radiotherapy for groing lesions may be considered.

Many varied forms of cancer immunotherapy

Just as the immune system is complex, then so are the many varied forms of cancer immunotherapy. One approach is to genetically modify a patient's own T-cells to make them target tumour cells. This so-called chimeric antigen receptor (CAR) therapy is a personalised form of cancer treatment. CAR T-cells have produced dramatic improvements when tested in clinical.

Other approaches have concentrated on lessening the natural inhibition of T-cells - in effect taking the "brakes" off so that the T-cells become potent killers free to destroy rogue cancer cells. This is done by producing targeted antibodies, known as monoclonal antibodies (MABs), that are directed against the "braking" molecules such as CTLA-4 and PD-1/PD-L1 (programmed death), which are known to act at T-cell inhibitors.

Immune checkpoint inhibitors (ICI)

Immune checkpoint inhibitors work by blocking checkpoint proteins that are made by some type of immune system cells, such as T cells, and some cancer cells. These checkpoints help keep immune responses from being too strong and sometimes keep T cells from killing cancer cells. When these checkpoints are blocked, T cells can kill cancer cells better. Examples of checkpoint proteins found on T cells or cancer cells include PD-1/PD-L1.

1. Borcoman et al. Annals of Oncology 30: 385-396, 2019.

5.1.3 Atezolizumab

A PD-L1 immune checkpoint inhibitor

A key component of the immune cellular pathway is PD-L1. Due to its presence in healthy cells, PD-L1 helps immune cells not attack healthy cells. Normally, the immune system fights foreign substances such as viruses and bacteria, but not its own healthy cells. Some cancer cells have a lot of PD-L1, which allows the cancer cells to "trick" the immune system into not attacking them as harmful foreign substances. Atezolizumab is an immunoglobular antibody designed to target the PD-L1 protein and block its interaction with its programmed receptors, thereby restoring T-cells, which are immune system cells that help protect the body from infection and may help fight cancer.

Atezolizumab is a humanised antiprogrammed death-ligand 1 (PD-L1) monoclonal antibody that inhibits PD-L1 and programmed death-1 (PD-1) and PD-L1 and B7-1 interactions, reinvigorating anticancer immunity.

5.1.4 The randomized OAK and POPLAR trials

Registration studies

To address this overarching question, it was suggested to analyze two patient cohorts, OAK (phase III) and POPLAR (phase II). These studies assess the results of different treatments focused on the inhibition of tumor growth in patients with non-small cell lung cancer. These patients received one of two treatments: chemotherapy (docetaxel) or immunotherapy (atezolizumab).

> OAK and POPLAR were randomized trials of atezolizumab 1200 mg intravenous (IV) every 3 weeks versus docetaxel 75 mg/m^2 IV every 3 weeks for patients with advanced NSCLC whose disease progressed on platinum-containing therapy.

The OAK study, a phase III trial of a treatment targeting the PD-L1 protein, is aimed at the treatment of advanced non-small cell lung cancer and showed that:

- Atezolizumab results in a substantial improvement in treating advanced non-small cell lung cancer over docetaxel, also these improvements are longer lasting.
- The proportion of patients with an objective response in the intention-to-treat (ITT) population did not improve with atezolizumab compared to docetaxel.
- There is an apparent discordance between progression-free survival and overall survival that could be explained by the initial increase in tumor volume due to increased immune infiltration, delayed anti-tumor activity or anti-tumor immune activation.

In the POPLAR study, aimed at the treatment of previously treated non-small cell lung cancer, it was found that:

- For patients with tumors containing high levels of PD-L1, they gained a greater benefit from atezolizumab.
- For patients with a low level of PD-L1 there is also a benefit to treatment with atezolizumab, this should be justified with further research to better understand the responses of these patients. One hypothesis is that atezolizumab increases cancer immunity through enhanced preparation of new cancer immune responses.

Therefore, the OAK study concludes that there is a greater survival benefit in the treatment of non-small cell lung cancer for patients previously treated with atezolizumab versus docetaxel, with a higher safety profile. The POPLAR study concludes that atezolizumab provides a survival benefit in previously treated NSCLC patients, and that PD-L1 expression on tumor cells (TC) and tumor-infiltrating immune cells (IC) is predictive of this benefit.

ImmuunPRO Hackathon

Personalized healthcare: How immunotherapy for non-small cell lung carcinoma can be tailored to individual patients

An exploration, analysis and modelling of patient data from the OAK & POPLAR clinical trial cohorts

PHC CATALYST

This report is developed by the members of the PHC Catalyst Alliance and has been made possible by a financial contribution of Roche Nederland B.V.

Roche project team

Rob Bohte
Mariëlle Gallegos-Ruiz
Cor Sibum
Project Sponsor: Judith van Schaik

Personalized Health Care Alliance members

Ewoudt van de Garde
Michel van den Heuvel
Daan Hurkmans

Accenture authors

Laury van Bedaf
laury.van.bedaf@accenture.com

Ana Faraon
ana.faraon@accenture.com

Thijs van den Broek
thijs.van.den.broek@accenture.com

Gaston Besanson
gaston.besanson@accenture.com

List of abbreviations

ABT
Analytical base table

AI
Artificial Intelligence

AUC
Area under the ROC curve

CIT
Cancer Immunotherapy

C-STATISTIC
Concordance statistic (equal to AUC)

EDA
Exploratory data analysis

EMA
European Medicines Agency

GBM
Gradient boosting machine

IC
Immune cells

NSCLC
Non-small cell lung carcinoma

OS
Overall survival

PHC
Personalized Health Care

RF
Random forest

ROC curve
Receiver-operator characteristic curve

SME
Subject matter expert

TC
Tumor cells

XGBoost
Extreme gradient boosting

1 Executive summary

The Dutch Personalized Health Care Catalyst Alliance and the cancer immunotherapy hackathon

Developments in science, society & technology have paved the way for a shift in thinking in many areas of life, also in the medical world: personalization. Possibilities, and increasingly expectations, of treatment have evolved from 'one-size-fits-all' to focusing on the differences between individuals and on the traits that uniquely characterize a patient. This approach is the basis for personalized healthcare: the acknowledgement that no two individuals are the exact same. By including a patient's characteristics such as their genetic make-up and the specifics of their disease, we can provide tailored treatments which optimize the outcome for each individual.

Since 2018 the Dutch Personalized Healthcare (PHC) Catalyst Alliance is active in increasing the understanding of the importance, opportunities, and challenges in the Personalized Healthcare space. The PHC Alliance is committed to accelerating the transition to personalized care by creating a receptive environment (sustainable, agile and adaptive health care system) in which optimal use is made of all available data, tools and knowledge (including best practices) and in which innovations that enable personalization quickly reach daily practice. The initiative for this development comes from Roche, however it was intended from the beginning that this Alliance would not be something for Roche, but rather made possible by Roche. Anyone who would feel compelled to join the movement is welcome. The Dutch Alliance is based on the principle of 'Combinatorial Innovation' (creating new value by combining knowledge, ideas and networks) which has led to a steady growth in 'coalition of the willing' parties that are directly or indirectly involved or want to contribute to the transition to a new healthcare system.

One of the in early projects the Alliance initiated was a data hackathon in which Roche NL, members of the Dutch PHC Alliance and Accenture NL collaborated. In this hackathon, medical expertise and Artificial Intelligence (AI) are leveraged to predict survival in Non-Small Cell Lung Carcinoma (NSCLC) patients treated with either immunotherapy or chemotherapy and to assess the effect of antibiotics on survival.

Immunotherapy is a type of treatment that targets the immune system and triggers it to fight disease, in this case cancer. By combining different information about the specific traits of a patient with the help of machine learning (ML) and AI, it is possible to assess treatment outcomes for each patient individually. Models that accurately predict which patients could benefit from immunotherapy may inform treatment choices and contribute to a higher quality of life.

Research questions address survival endpoints prediction and antibiotics effect

In this research project, we worked together with health care professionals from academic expertise centres - all Alliance members - to refine and prioritize research questions to be analyzed. Through a series of workshops, the following research questions were selected:

- Can we predict which patients survive after 2 years?
- Can we predict who has no cancer progression after 3 months of treatment?
- Which predictor variables are correlated with overall survival after 2 years and with progression-free survival at three months?
- What is the effect of antibiotics on the OS at 2 years?

To predict survival, we leveraged the power of machine learning that allows us to find new correlations within the datasets. First, we conducted an exploratory data analysis (EDA) to measure correlations between variables and to define the targets for overall survival at 2 years and progression-free survival at 3 months. After preprocessing the data, different modelling strategies were tested in combination with algorithms such as random forest (RF), gradient boosting machine (GBM), and extreme gradient boosting (XGBoost).

Table 1: Model results for overall survival predictive model

AUC for OS at 2 years	Sensitivity	Specificity
0.81	0.67	0.76

The OS prediction model makes a binary classification, meaning that asking the question "is the patient alive at 2 years from the start of treatment?" has two possible answers (a patient is either alive or not at a given time). Predictive models for binary outcomes are evaluated using the Area Under the Receiver - Operator Characteristic (ROC) Curve, or AUC. Our model for OS at 2 years predicts survival with AUC 0.81 and a second predictive model, developed for PFS at 3 months, has the predictive accuracy AUC 0.64.

In addition to the models built, we conducted a survival analysis using the Kaplan-Meier estimator to detect the effect of antibiotics on survival. We found that patients who receive antibiotics have a significantly lower overall survival than those who are not given a course of antibiotics ($p = 0.01$).

Evaluation and next steps to realize personalized healthcare

The survival model generated in this project can be used to predict for each patient the probability of survival, proving the model potential to support clinicians and patients in their decisions. Thus, the important question regarding predictive models is: when is a model accurate enough for patients and for physicians? And how can it be applied in the clinic? Treatment options should be considered carefully based on both model performance and the patient's perspective. For this reason, it is advisable to test the model on real-world data and to include the patients themselves in the discussion on how to assess treatment options supplemented by AI.

2 Introduction: the importance of identifying which patients benefit from immunotherapy

A showcase for the PHC Alliance to accelerate the transition towards personalized healthcare

Personalized Health Care shifts the focus from the traditional 'average patient' resulting in trial-and-error treatments to recommendations tailored to patients based on their genomic profile and other individual characteristics[1]. These unique traits can be used in an early stage to detect disease, or later in the standard of care to make personalized predictions for disease progression and treatment results. Personalized care can yield better outcomes in patient stratification by ultimately allowing for subcategories of n=1 patient[2], as well as fewer side effects like drug toxicity. In the case of immunotherapy only a fraction of the patients has a lasting response to therapy[3], showing that a personalized approach is needed to find the right therapy for each patient. Thus, the multidisciplinary Dutch Personalized Healthcare (PHC) Catalyst Alliance was brought together to accelerate the transition towards a personalized healthcare model[4,5]. One of the early projects the Alliance initiated was meant to bring together Alliance members and Accenture to leverage medical expertise and AI in order to determine which Non-Small Cell Lung Carcinoma (NSCLC) patients will benefit from immunotherapy[4].

Studies show the benefits of immunotherapy for some NSCLC patients

To address this overarching question, it was suggested to analyze two patient cohorts, OAK[6] (phase III) and POPLAR[7] (phase II). These studies assess the results of different treatments focused on the inhibition of tumor growth in patients with non-small cell lung cancer. These patients received one of two treatments: chemotherapy (docetaxel) or immunotherapy (atezolizumab).

A key component of the immune cellular pathway is PD-L1. Due to its presence in healthy cells, PD-L1 helps immune cells not attack healthy cells. Normally, the immune system fights foreign substances such as viruses and bacteria, but not its own healthy cells. Some cancer cells have a lot of PD-L1, which allows the cancer cells to "trick" the immune system into not attacking them as harmful foreign substances. Atezolizumab is an immunoglobular antibody designed to target the PD-L1 protein and block its interaction with its programmed receptors, thereby restoring T-cells, which are immune system cells that help protect the body from infection and may help fight cancer.

The OAK study, a phase III trial of a treatment targeting the PD-L1 protein, is aimed at the treatment of advanced non-small cell lung cancer and showed that[6]:

- Atezolizumab results in a substantial improvement in treating advanced non-small cell lung cancer over docetaxel, also these improvements are longer lasting.
- The proportion of patients with an objective response in the intention-to-treat (ITT) population did not improve with atezolizumab compared to docetaxel.
- There is an apparent discordance between progression-free survival and overall survival that could be explained by the initial increase in tumor volume due to increased immune infiltration, delayed anti-tumor activity or anti-tumor immune activation.

In the POPLAR study, aimed at the treatment of previously treated non-small cell lung cancer, it was found that[7]:

- For patients with tumors containing high levels of PD-L1, they gained a greater benefit from atezolizumab.
- For patients with a low level of PD-L1 there is also a benefit to treatment with atezolizumab, this should be justified with further research to better understand the responses of these patients. One hypothesis is that atezolizumab increases cancer immunity through enhanced preparation of new cancer immune responses.

Therefore, the OAK study concludes that there is a greater survival benefit in the treatment of non-small cell lung cancer for patients previously treated with atezolizumab versus docetaxel, with a higher safety profile[6,8]. The POPLAR study concludes that atezolizumab provides a survival benefit in previously treated NSCLC patients, and that PD-L1 expression on tumor cells (TC) and tumor-infiltrating immune cells (IC) is predictive of this benefit[7].

During the journey to access data, the hackathon evolved into a proof of concept

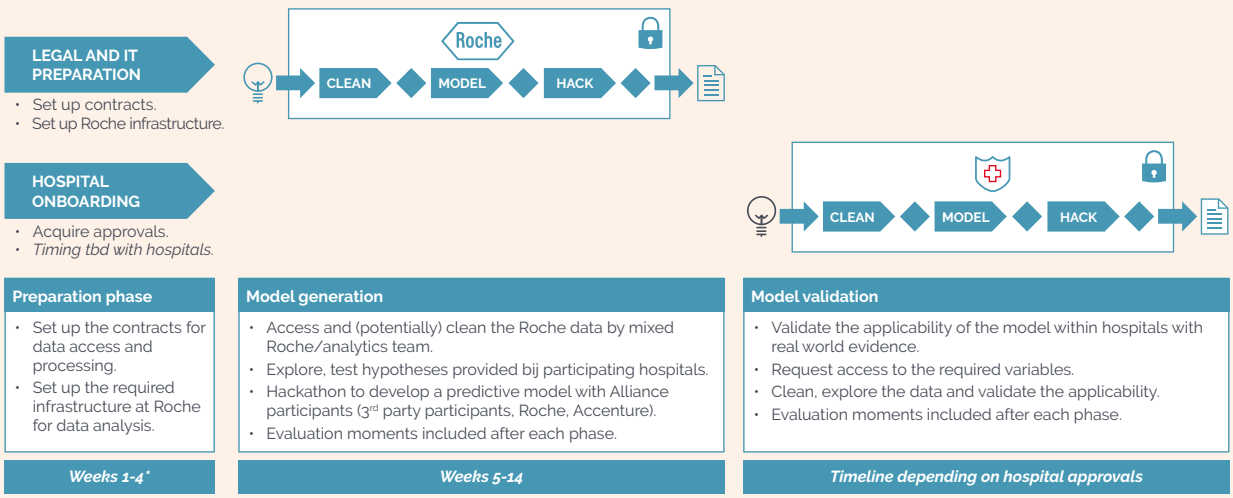
In order to access Roche clinical trial data, specific questions are needed with a valid scientific reason or aiming to improve patient care. To complete the data access request, the healthcare professionals, members of the PHC Alliance, have provided their initial research questions as follows:

- To determine the effect of concomitant medications on objective response and/or survival (PFS/OS) of patients treated with atezolizumab, and to determine its confounding factors.
- To determine patient baseline factors influencing survival.

The initial data hackathon proposal was to build a predictive model for cancer immunotherapy (CIT) in NSCLC patients. Due to the restricted nature of the data, it was not feasible to organize a public hackathon in the format originally envisioned. However, the journey to find and access data resulted in an innovative setup: PHC Alliance members lend their expertise in the analysis and receive insights, while the clinical trial data remains in a virtual data room. Therefore, it was decided that a prediction model based upon the POPLAR and OAK studies will be validated for each of the 3 hospitals separately.

Once the data access request was approved, the format of the analysis was updated to contain three phases as illustrated below: a preparation phase to set up legal agreements & infrastructure required for analysis, a phase to generate a machine learning predictive model for NSCLC patients, and a phase to validate the model on real world data (RWD) from hospital patients.

Figure 1: Data analysis extended planning



For details on the data access timelines, see Appendix A.
The preparation phase was completed successfully when Roche granted clearance for data access to Roche NL and the Accenture AI team. Subsequently, data from two study cohorts (OAK and POPLAR) were made available in a virtual data room in the form of SAS datafile extracts. Marking the start of the model generation phase, a pipeline was built to load, explore and analyze the data using R in the RStudio environment.

3 Pursuing the analytical workflow together with Alliance members: methods and results

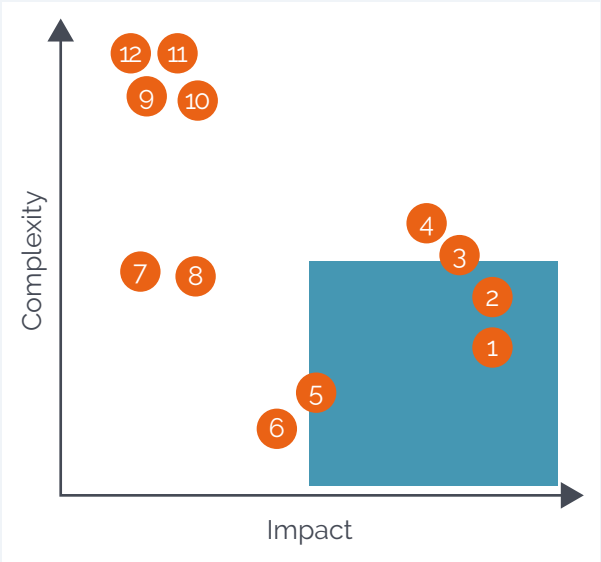
a. Workshops

Following an initial exploration of the data, the research scope and analysis strategy were developed and validated together with selected Alliance members in a series of interviews and ideation workshops.

In the first workshop, together with domain experts, we assessed existing questions & prioritized new research questions in more detail. Fig.2 shows the full list of these questions, mapped according to impact and effort. Based on this mapping, the following research questions were selected for the current project scope:

- Can we predict which patients survive after 2 years?
- Can we predict who has no cancer progression after 3 months of treatment?
- What is the effect of antibiotics on the disease progression?
- Which predictor variables are correlated with overall survival after 2 years and with progression-free survival at three months?

Figure 2: Prioritization of research questions



1. **Predict PFS at 3 months from baseline**
2. **Predict OS at 2 years from baseline**
3. Can we predict the duration of the response?
4. Are there routine test panels in the lab and how often are they applied? Can we get earlier warning on progression or response based on panel results?
5. **What is the effect of antibiotics on PFS and OS? And how do these differ between the arms?**
6. **Direct correlation of all variables with OS & PFS**
7. Correlation between QoL & survival
8. What disease/medication do patients have at baseline? And the relation on PFS
9. Predict QoL when patient survives after 2 years
10. Find important dissimilarities or similarities between patients
11. Feasibility exercise for missing data
12. Can we do a time analysis on the lab values?

Within this scope, we selected for further analysis datasets that hold data about subject level analysis, concomitant medication, treatment response, laboratory test results, medical history, and vital signs. The overview of the OAK/POPLAR datasets used is available below: the subject level analysis contains clinical measurements, treatment arm, study completion information, PD-L1 expression on tumor cells and on immune cells, mutations and other patient characteristics. Comedication data contains the medication class and dose, including antibiotics baseline data. Other datasets with predictive variables are the lab values, vital signs, and medical history. Lastly, the response data is used to identify which patients are alive and have no cancer progression and thus define targets for the predictive models.

Table 2: Data sources

Subject level analysis
Age, sex, race, ECOG score
Metastases (nr, flag per organ)
Histology
PD-L1 on TC & IC cells
Mutation KRAS, EGFR, ALK
Alive flag,
Study completion
Treatment completion
Therapy line
Tobacco use history
Time of exposure
Time of diagnostic

Comedication
Comedication name, class, dose, antibiotics

Vital signs
Vital signs test name, visits name, results

Medical history
Event category
Number of events, epochs

Lab values
Test parameters, lab values with ranges, lab tests with flags

Treatment response
Response metrics, response time, tumor progression

To address the research questions above, we pursued a modelling strategy starting with the exploratory data analysis (EDA) of the tables in scope including definition of the OS and PFS targets. This was followed by feature engineering, normalization of the data and building of the analytical base table (ABT). We presented this in the second workshop, together with the analytics approach illustrated in Appendix B for the supervised machine learning models for OS and PFS. At this time, a similar survival analysis in NSCLC patients was published by Hopkins *et al.*[9]. Based on 4 cohorts including OAK & POPLAR, that team developed a prognostic tool for patient risk stratification. Therefore, we assessed our data analysis & methodology against the benchmark provided by this prognostic model. To do so, we reviewed the research paper together with the Alliance members. This research is summarized in Fig.3, and it is worth noting that most of the important variables published by Hopkins *et al.* had been selected during our first workshop in order to be included in the predictive model.

Figure 3: Summary of research by Hopkins *et al.*[9]

Hopkins *et al.* summary

Setup
Build a model on the atezolizumab arm of OAK/ POPLAR cohorts. Test on additional cohorts incl. docetaxel arm.

Create prognostic groups for OS & PFS in NSCLC patients initiating atezolizumab.

Data

Important variables for the Hopkins *et al.* model are similar to parameters for baseline predictions selected in the current analysis.

Approach

Survival analysis is conducted to make a decision support tool. Additionally, the random survival forest method is explored.

Outcome

The resulting pre-treatment prognostic tool can be used to evaluate patient risk groups from low to high (c-statistic = 0.76)

Prognostic tool outperforms existing prognostic scoring tools.

Strongest predictors of OS at 2 years: CRP, LDH, dNLR, albumin, ECOG performance score, time since metastatic diagnosis, nr. of metastases.

Comparison of pooled randomised arms of OAK/ POPLAR indicates that benefit of CIT over chemo-therapy is greater in patients in the lower risk group.

Predictor variables at the baseline

1. Sex
2. Age
3. Race
4. Body Mass index
5. Tobacco use history
6. ECOG performance score
7. Histology
8. Stage
9. Time since diagnosis
10. Nr. of prior treatments
11. PD-L1 expression
12. Nr. of metastases
13. Neutrophil to lymphocyte ratio (NLR)
14. Derived NLR
15. Lymphocite to monocite ratio (LMR)
16. Platelet to lymphocite ratio (PLR)
17. Eosinophils
18. Lactate dehydrogenase (LDH)
19. C-reactive protein (CRP)
20. Alkaline phosphatase (ALP)
21. Calcium
22. Haemoglobim
23. Albumin

In the third workshop, the prediction models we built for OS and PFS were reviewed and evaluated to define possible applicability in the clinic. The important variables for these models were compared with results from the correlation analysis that was done to identify which predictor variables are highly correlated with the targets or with other variables. By showing that variables are consistent between the predictive model and the correlation analysis, we increase the interpretability of the model and, with it, we demonstrate that the need for explainable decision support tools in the clinic can be satisfied with the help of machine learning and AI.

In the sections below, the research questions on survival prediction and correlation analysis are grouped together since they relate to a supervised machine learning problem. The antibiotic analysis is treated separately given that it is based on a survival analysis (Kaplan-Meier estimator) that fits better within a statistical framework.

b. Machine learning models

Exploratory data analysis

The OAK and POPLAR variables were mapped onto each other to combine datasets. Where variable names did not match between cohorts, the mapping was done with the support of the Roche data management team. Based on the research questions, input received during workshops from domain experts, and data exploration, a selection of variables was made for the analytical base table (ABT) for the machine learning model. This variable selection can be found in Appendix C and the treatment for missing values is found in this appendix. To generate the predictive model, we use an ABT that contains the anonymized patient identifiers, the target endpoints and the engineered features. First, unique patient IDs were used to build the ABT: each row represents a distinct patient, meaning the ABT is unique at patient level (n = 1512 patients).

Endpoint definition

The population survival endpoints were defined for the 1512 patients in the combined OAK and POPLAR cohorts. Three groups were defined for OS at 2 years:

- Group 1: Patients for whom we know they are not alive and did die.
- Group 2: Patients for whom we know they are alive and did not die.
- Group 3: Patients that are not recorded to die but whose last recorded alive date is below 730 days. This could be due to censoring (for example lost in follow-up, withdrawal etc.) or data errors. For the prediction model only Group 1 and Group 2 are in scope.

Table 3: Target definition for overall survival at 2 years

OS groups	In scope	Nr. of patients
Group 1	Yes	986
Group 2	Yes	393
Group 3	No	120

To define PFS at 3 months, we analysed the patient's alive status and tumor progression. We notice that many patients have the analysis study day (variable ADY) recorded around the 12 week visit. The average day for the 12 weeks tumor assessment is 85.1, therefore day 85 is used to count patients alive at 3 months. For progression-free status, a patient's last tumor assessment by an investigator must not indicate progressive disease (PD). Additionally, when the progressive disease is analysed by an investigator, the outcome is negative. Based on these criteria 3 groups are identified.

- Group 1: patients with progressive tumor before 3 months.
- Group 2: patients who are alive at 3 months and have progression-free status.
- Group 3: patients who do not have a 12 week assessment but also no death date recorded. Since there is no 12 weeks visit or death record available for these patients, it cannot be concluded which of the first two group2 they would belong to, if any. Therefore, for the prediction of PFS, only the first two groups are in scope.

Table 4: Target definition for progression-free survival at 3 months

PFS groups	In scope	Nr. of patients
Group 1	Yes	637
Group 2	Yes	770
Group 3	No	105

Feature engineering

Using the input variables, we prepared the data in a format that can be interpreted by the machine learning algorithm. This preparation entails data normalization, feature engineering, with some of these data transformations shown in Table 5. For example, age is a continuous variable and can be used as is. Conversely, histology information has the categorical values 'squamous' or 'non-squamous' and cannot be interpreted by an algorithm as such. To address this, a method called 'one-hot encoding' is applied where categorical variables are converted into binary ones. Here, histology is transformed into a new variable histology squamous with values 1 or 0 corresponding to yes and no, respectively. While the values 'squamous' or 'non-squamous' cannot be used by the algorithm to train a model, values 1 and 0 can be used for this purpose. Another example is the Eastern Cooperative Oncology Group (ECOG) performance score[10] which is used to assess the daily living abilities of patients on a scale from 0 (asymptomatic) to 5 (dead). Based on the exclusion criteria only patients with ECOG scores 0 or 1 were included in the OAK & POPLAR cohorts. Within our programming environment, these values were initially interpreted as factors meaning the ECOG score could be one of two categories, without a relationship between them. In real life however this is not the case since we know there is an order between the two scores, namely patients with ECOG 0 are less hindered in their daily activities than patients with ECOG 1. To translate this information into code, the ECOG variable was converted from a factor into a numerical variable. Following a similar analysis, all the variables used as model predictors are transformed into new features where needed.

Table 5: Feature engineering based on variables from the data sources

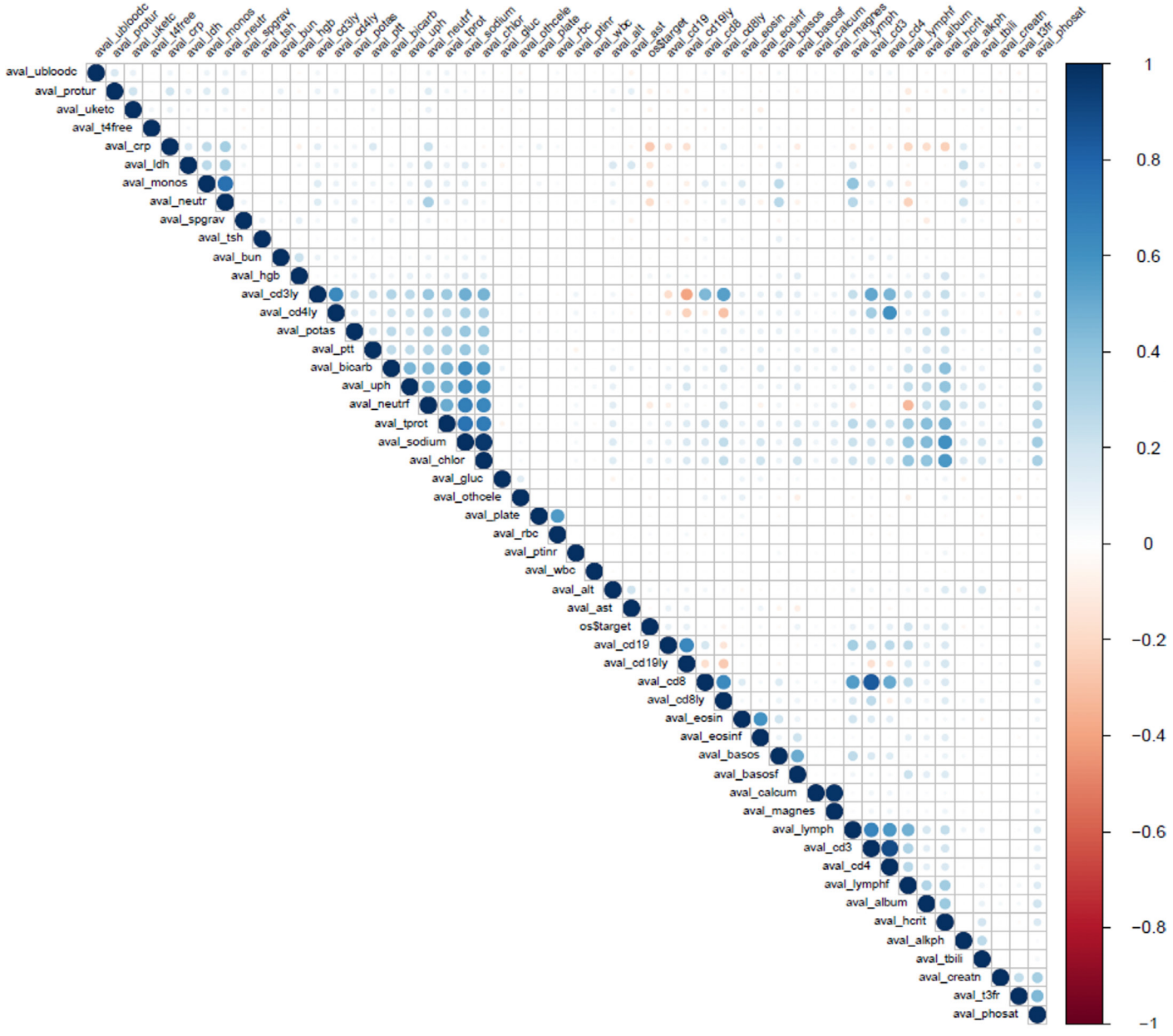
Variable	Treatment for feature engineering	Variable	Treatment for feature engineering
Age	None (num ric)	Nr. of prior treatments	None (numeric)
Sex	One-hot encoding (example: 'male' or 'female' becomes 0 or 1)	EGFR mutations	One-hot encoding
Histology	One-hot encoding (example: 'squamous' or 'non-squamous' becomes 0 or 1)	KRAS mutations	One-hot encoding
Raw PD-L1 score	None (numeric)	ALK mutations	One-hot encoding
PD-L1 score	Encoding (0, 1, 2, or 3)	Tobacco use history	One-hot encoding
ECOG performance score	Factor encoding as numeric (0 or 1)	Baseline protein levels	Result at baseline, reference, grade
BMI	None (numeric)	Comorbidity	Present at baseline yes/no, years in the past, amount of years
Weight	None (numeric)	Comedication	Taken at baseline; yes/no flag and dose
Stage	One-hot encoding	QoL at baseline	Score at baseline
Time since diagnosis	None (numeric)	Response parameters	One-hot encoding (example: Param_last_tum_assess_yes is 0 or 1)

Correlation analysis & important features

In this phase a correlation analysis was conducted to assess interactions amongst features and between features and target (OS and PFS, respectively). Pre-processing included removal of features with near zero variance because these are not informative for modelling, and Spearman correlations were used on complete observations (each observation is one row). This resulted in 1407 observations and 441 features for PFS and 1379 observations and 437 features for OS. Hierarchically clustered correlation plots are available for each ABT component: basic patient characteristics, comedication, lab values, medical history, and vital signs.

The correlations for lab values and OS may be seen in Fig.4, where aval_ refers to analysis value. Here we may see there is some correlation between CD3, CD4; neutrophils, total protein, sodium, chloride, and CD19 are also correlated. Overall there is no significant correlation between features and the OS target.

Figure 4. Hierarchical clustering of correlations between lab values and OS target



The correlation analysis informs the selection of features for the machine learning model and adds interpretability; complete results for this analysis can be found in Appendix G. In our analysis, the algorithms used (GBM and XGBoost) are not affected by feature multicollinearity. For both OS and PFS targets, none of the features is highly correlated with the target. Some of the features with higher correlation with OS are C-reactive protein (CRP) analysis value, CRP high (flag), nr. of metastases, grade I hemoglobin, ECOG performance score, neutrophils high (flag), hemoglobin normal (flag), lymphocytes value, and hematocrit normal (flag). For PFS, some variables with higher correlation are CRP analysis value, nr. of metastases, grade I hemoglobin, neutrophils high, ALK phosphatase, albumin, lactate dehydrogenase high (flag), liver metastasis (flag), lymphocytes value, CD9, CD4, total protein, and TC/IC mean. Some of these variables are also important predictors for the models we built to predict OS and PFS. Table 6 shows the variables importance of the best performing OS model based on the XGBoost algorithm with elastic net feature selection, and the variable importance of the best performing PFS model based on the GBM algorithm and elastic net feature selection. Additionally, these variables were also found to be important in the Hopkins *et al.*[9] model (Fig.3). A description of the lab tests is included in Appendix C.

Table 6: Feature importance for the OS and PFS predictive models

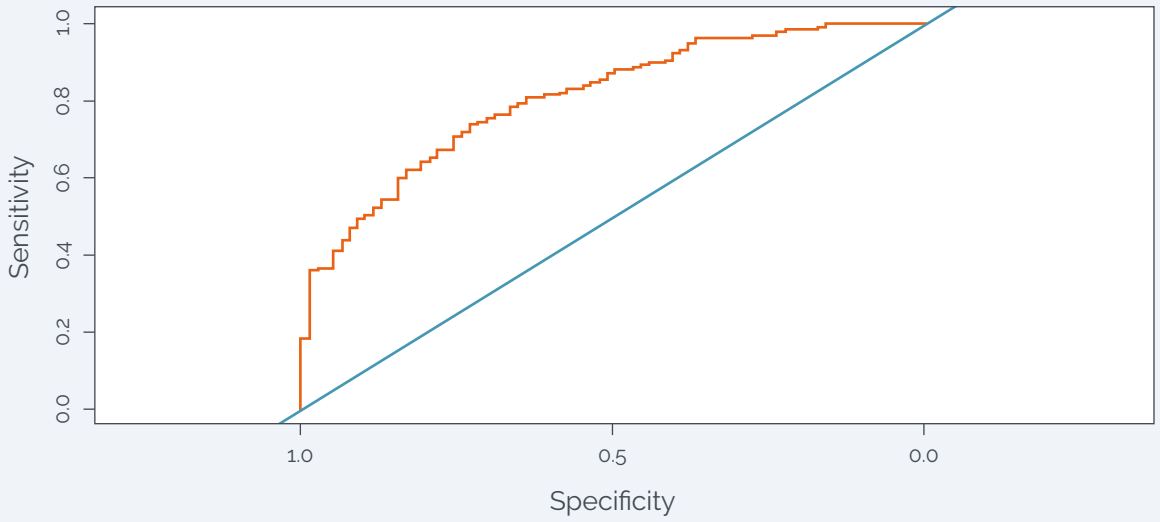
Nr.	Variables OS Model	Nr.	Variables PPS Model
1	aval_crp	1	aval_crp
2	metsites	2	time_since_diagnosis
3	time_since_diagnosis	3	aval_cd19
4	aval_neutr	4	aval_neutr
5	lbnrind_crp_HIGH	5	metsites
6	tc_ic_mean	6	lbnrind_album_NORMAL
7	lbnrind_hgb_LOW	7	aval_lymphf
8	lbnrind_t3fr_NORMAL	8	aval_f8915070
9	ecoggr	9	kras_neg_alk_neg
10	aval_f8915091	10	ecoggr
11	aval_chlor	11	aval_f8915040
12	lbnrindidh_HIGH	12	PULSE
13	aval_f8915094	13	lbnrindldh_HIGH
14	aval_ldh	14	aval_cd19ly
15	aval_lymphf	15	gradel_hgb
16	treatment_atezolizumab		
17	met_liver_yes		
18	met_peff_yes		
19	aval_f8915043		
20	lbnrind_cd8ly_unknown		
21	aval_alt		
22	aval_sodium		
23	cmclas_flag_opioid_analgesics		
24	mh_ongoing_prim_sys_org_class_general_disorders_and_administration_site_conditions		
25	hist_squamous		
26	aval_spgrav		
27	mh_resolved_prim_sys_org_class_respiratory_thoracic and mediastinal disorders		

Predictive modelling results & evaluation

In the supervised ML modelling phase, different algorithms were tested including random forest (RF), gradient boosting machine (GBM) and XGBoost (extreme gradient boosting). For more details see the modelling strategy in Appendix B.

Clinical models for survival endpoints are commonly evaluated with the concordance statistic (c-statistic). For a given survival endpoint the assessment can be also interpreted as a binary classification problem, meaning that asking the question "is the patient alive at 2 years from the start of treatment?" yields one of two answers, a patient is either alive or not at that time. Predictive models for binary outcomes can be evaluated using the Area Under the Receiver - Operator Characteristic (ROC) Curve, or AUC[11]. For binary outcomes, the AUC is equivalent to the c-statistic[12] and accounts for the whole performance of the model, since each point on the ROC curve indicates the model sensitivity and specificity for that threshold. Sensitivity indicates the percentage of patients correctly predicted by the model to be alive at 2 years and specificity is the percentage of patients correctly predicted not to be alive.

Figure 5: ROC curve of the OS model



The model reaches a performance of AUC = 0.81 using the XGBoost algorithm and feature selection with elastic net. The sensitivity for this model is 0.67 and specificity is 0.76, at a threshold of 0.35. For a discussion on setting the model threshold, see Appendix D. This OS model has higher performance than the tool developed by Hopkins *et al.*, which has a c-statistic of 0.72 on the development cohort and 0.76 on the validation cohort[9]. The model for PFS at 3 months uses a gradient boosting machine (GBM) algorithm and feature selection with elastic net. For this model, AUC is 0.64, sensitivity 0.77 and specificity 0.33 at the 0.5 threshold. The report for this model is available in Appendix E.

Table 7: Summary of ML models built for OS prediction

Nr.	OS Model description	AUC	Sensitivity 0.5	Specificity 0.5	Kappa 0.5	Sensitivity 0.35	Specificity 0.35	Kappa 0.35	# Patients survive, predicted correctly at 0.35	# Patients survive, predicted incorrectly at 0.35	Nr. of features used
1	Random forest without f variables	0.81	0.32	0.96	0.35	0.59	0.82	0.4	46	32	42
2	XGBoost without the f-variables	0.78	0.41	0.9	0.35	0.58	0.77	0.34	45	33	31
3	Ensemble model	0.61	0.15	0.95	0.14	0.27	0.89	0.189	21	57	130
4	Gradient boosting machine on small set	0.77	0.1	0.99	0.13	0.54	0.85	0.4	42	36	5
5	Random forest with all features	0.83	0.2	0.97	0.24	0.59	0.83	0.42	46	32	58
6	Gradient boosting machine trained on sens feature selection with elastic net with vital signs	0.80	0.22	0.98	0.25	0.56	0.84	0.41	44	34	17
7	XGBoost without vital signs feature selection with elastic net	0.81	0.41	0.9	0.35	0.67	0.76	0.41	53	25	31
8	Gradient boosting machine without vital signs feature selection with elastic net	0.80	0.22	0.97	0.24	0.58	0.86	0.44	45	33	17
9	Random forest without vital signs feature selection with elastic net	0.81	0.23	0.97	0.26	0.56	0.83	0.4	45	33	55
10	XGBoost feature selection with trees	0.81	0.3	0.96	0.31	0.65	0.77	0.39	51	27	29
11	Gradient boosting machine feature selection with trees	0.79	0.17	0.98	0.19	0.59	0.85	0.44	46	32	20
12	Random forest feature selection with trees	0.81	0.17	0.97	0.18	0.64	0.84	0.48	50	28	82
13	First model: Gradient boosting feature selection based on variable importance	0.78	0.28	0.96	0.3	0.58	0.78	0.35	45	33	14

Table 8: Summary of ML models built for PFS prediction

Nr.	OS Model description	AUC	Sensitivity 0.5	Specificity 0.5	Kappa 0.5	Sensitivity 0.35	Specificity 0.35	Kappa 0.35	# Patients survive, predicted correctly at 0.5	# Patients survive, predicted incorrectly at 0.5	Nr. of features used
1	XGBoost with elasticnet feature selection	0.61	0.65	0.44	0.09	0.87	0.25	0.10	101	53	19
2	Gbm with elasticnet feature selection	0.64	0.77	0.33	0.10	0.94	0.09	0.04	120	34	15
3	Rf with elasticnet feature selection	0.63	0.70	0.44	0.15	0.90	0.25	0.17	109	45	23
4	XGBoost tree without f variables, vital signs	0.56	0.66	0.41	0.02	0.86	0.22	0.09	108	52	45
5	XGBoost without f variables, vital signs	0.59	0.72	0.42	0.15	0.12	0.19	0.12	111	43	78
6	Rf without vital signs	0.62	0.73	0.43	0.17	0.95	0.19	0.08	113	41	82
7	GBM without vital signs	0.64	0.82	0.48	0.19	0.97	0.06	0.04	120	34	22

Performance evaluation by treatment arm

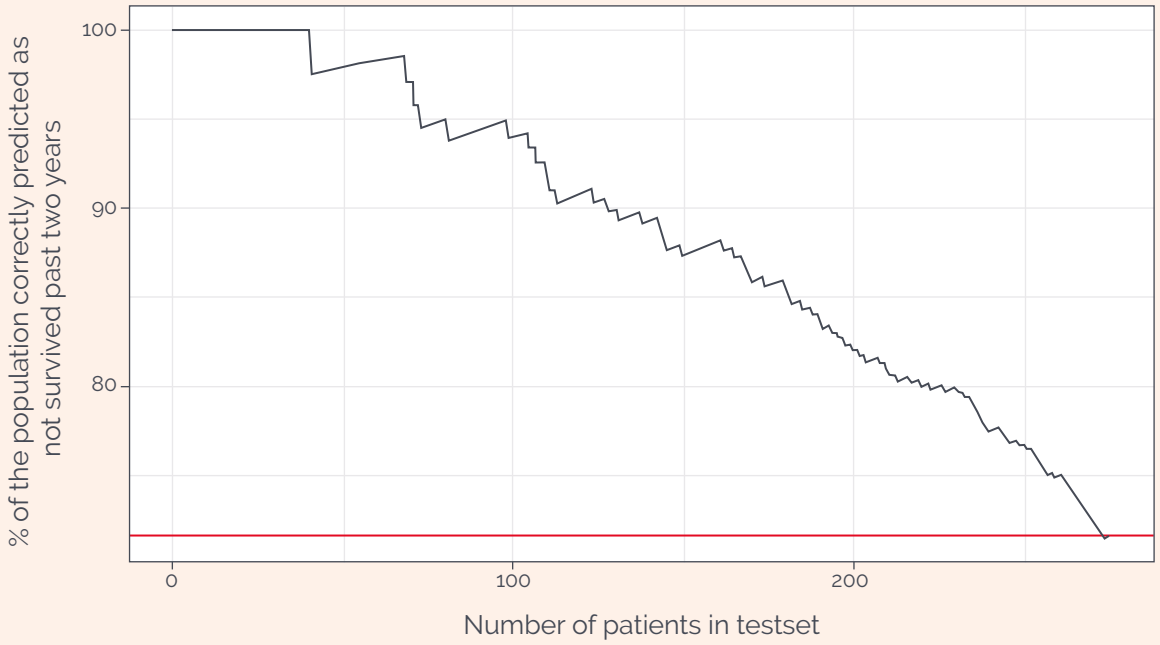
In the modelling design, OAK & POPLAR patient data was used to train the OS model irrespective of the treatment arm. To test whether more patients in the atezolizumab group are predicted to die when in fact they survive and to test prediction accuracy, the cohort was split by treatment arm. For a threshold of 0.5, the model makes false classifications for 18.4% of the patients in the atezolizumab group and for 14% of patients in the docetaxel group.

When the threshold is moved to 0.35, this gap decreases: 8% of the atezolizumab patients are falsely classified as not surviving, while for the docetaxel group the figure is 8.67%. At the same threshold we analyzed prediction accuracy per treatment arm and found the accuracy of the docetaxel group (76%) to be a bit better than that of the atezolizumab group (70%). This difference is small and mainly driven by the difference in the patients that do not survive and where the model is predicting they do. As this is opposite to our hypothesis, we do not have a reason to assume there is a significant difference in the OS model performance between the two treatment arms.

Performance evaluation for clinical applicability

The clinical relevance of the OS prediction model we developed is to predict whether a patient is alive at 2 years or not. Nevertheless, it may be just as important to reframe this by asking what is the chance that a patient will most likely die within a given period. To assess this, we ranked patients in a hit rate curve (Fig.6) by the probability they will not survive after 2 years. In the test set below, about a third of the patients (n=100) have a 95% chance of not being alive compared to the combined OAK/POPLAR cohort where patients have a 70% probability. Knowing with a high probability that one might die within a timeframe may enable patients to make different choices regarding their treatment and quality of life.

Figure 6: Patients ranked by probability they will not survive after 2 years



c. Antibiotic analysis

For the analysis of survival outcomes, it was hypothesized that antibiotics had a negative effect on overall survival especially in the atezolizumab treatment group. This is in line with research by Chalabi *et al.*[13], who performed a pooled ad-hoc analysis on the OAK and POPLAR trail data and suggested that use of antibiotics (ABT) from 30 days before treatment until 30 days after start of treatment is associated with worse outcomes.

On the next page is the table with the classes of drugs that are labeled as ABT in this analysis. Next to the classes are the number of patients that have taken this drug in the time window of 30 days before treatment start until 30 days after treatment start. In total 170 patients in the atezolizumab treatment group received ABT and 207 patients in the docetaxel treatment group received ABT.

Table 9: Antibiotic administration per NSCLC treatment arm

Drug class	Nr. patients in docetaxel arm	Nr. patients in atezolizumab arm
quinolone antibiotics	107	75
penicillin	69	71
cephalosporin antibiotics	55	35
macrolide antibiotics	26	26
carbapenem antibiotics	7	20
glycopeptide antibiotics	13	15
aminoglycoside antimicrobials	13	6
sulfonamides	4	11
lincomycin antibiotics	8	4
tetracyclines	8	6
miscellaneous antimicrobials	7	3
oxazolidinone antibiotics	NA	5

To analyze the effect of ABT on overall survival, Kaplan-Meier plots are made and with the log-ranking test we tested if the difference between the survival curves is significant. Overall survival is defined as the last date a patient was known to be alive, with the event being if the patient has died (yes or no). If the alive state of a patients was unknown, the patient was listed at censored. The Kaplan-Meier plot is shown in Fig.7; for the analysis report see Appendix F.

Fig.7 shows that patients in the pink group, who were receiving atezolizumab as treatment and no antibiotics, had the best survival chances, while patients in the blue group who received atezolizumab in combination with antibiotics had a lower survival rate. The difference between these survival curves is significant, meaning that receiving ABT around the start of treatment is negatively impacting the survival outcome. This result, which is in line with the Chalabi et.al. findings, is potentially very relevant because choices for use of antibiotics during CIT could be adjusted to better meet patient needs.

As is often the case, these results trigger additional research questions where further investigation may lead to clinically actionable insights. The main questions are:

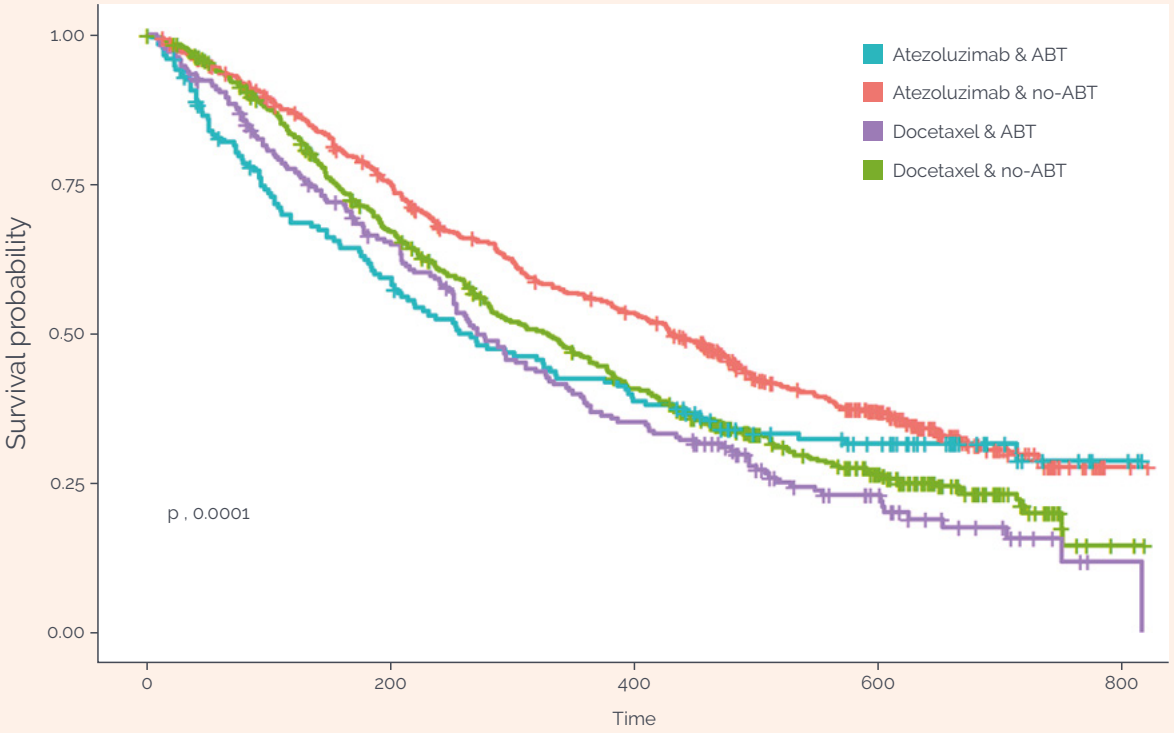
- By varying the timepoints when antibiotics are used, what is the effect on overall survival?

The hypothesis is that antibiotics will have a negative effect on survival not just in the -30 to 30 days window used by Chalabi et al., but also at different timepoints. If this is the case, this knowledge becomes highly relevant for the clinic since the duration of the antibiotics course or the class of antibiotics can be altered.

- If the time window between randomisation and start of the treatment increases, is there a negative effect on overall survival?

The hypothesis is that if a patient is ill at randomisation then treatment will be postponed, meaning that patients with longer time between randomisation and the start of treatment may be more ill. If this is the case, the length of this window can be used as an indicator for lower overall survival.

Figure 7: Kaplan-Meier survival plot showing stratification per treatment arm and antibiotic administration



4 Applicability on real-world data needs responsible and explainable AI

To accelerate the use of personalized healthcare in the Netherlands, a series of milestones need to be reached. The current project serves as inspiration for how to use AI to make a clinically meaningful prediction model and treatment tool, it also yields learnings around the process to access and share data in a novel, multidisciplinary ecosystem. To leverage these results further, the model needs to be tested on real world data (RWD) from hospitals. Validation of a predictive model on RWD is necessary for the model to be adopted in the clinic, and when such a decision support tool is used by patients and physicians, the goal of personalized healthcare draws nearer.

Figure 8: Milestones needed to accelerate personalized healthcare

To pilot the current model in hospitals, a roadmap is needed to map the necessary actions. Based on current project learnings, the following are needed to validate the model in hospitals:

- Assess which features of the current model are available in the hospitals
- Clearly define the validation criteria to identify when the pilot is considered successful
- Define when is the model performance acceptable and trade-offs (for example loss in prediction accuracy vs. increased explainability)
- Define explainability criteria (for example, correlation analysis or reduced number of predictor variables)
- Scope the groups of patients for whom the model could be used
- Assess whether there is potential for bias in the way this model is built and applied, and what factors could mitigate it

Using machine learning and AI to support treatment decisions has important consequences for all the parties involved - patients, healthcare practitioners, and others. Therefore, it is increasingly necessary to ensure AI tools are used to help patients in a fair and responsible manner. With the help of PHC Alliance members, we have identified the following risks for bias:

- Clinical trial exclusion criteria (the model development cohorts included only patients with ECOG performance score 0 or 1)
- Potential language barriers
- Potentially the patients who enroll in clinical trials may be more proactive about their health, this may be correlated with socio-economic differences between patient groups
- Current lack of patient perspective on the use of AI in treatment planning

An assessment of the model is performed to detect potential bias

While the framework for responsible AI and explainable AI is much wider, we hope to spark a discussion on practicing responsible AI when collecting data and building predictive models. To do this in our project, we started by assessing the inclusion and exclusion criteria of both OAK and POPLAR studies. According to the inclusion criteria, patients with ECOG score 0 or 1, older than 18 years and with locally advanced or metastatic NSCLC were included in the studies. The exclusion criteria were cancer-specific (active or untreated central nervous system metastasis), general medical exclusions (recent pneumonitis), criteria related to docetaxel (prior docetaxel treatment), or criteria related to atezolizumab (history of autoimmune disease, CD137 agonists, anti- CTLA4, anti-PD-L1, or anti-PD-1 therapeutic antibodies, or PD-L1-PD-1 pathway-targeting agents). Next, we assessed whether there are sensitive groups that require special consideration - these groups may typically be based on race, sex, age, or socio-economic status. For a model to be applied successfully in real life, it needs to be developed on cohorts that are representative of the wider population where the model will be used.

One of the factors that may impact applicability are ethnic factors, defined by the European Medicine Agency (EMA) as factors relating to the genetic, physiologic, cultural and environmental characteristics of a population[14]. Notably, the OAK study (phase III) already satisfies a standard EMA requirement to enroll patients of different ethnicities in phase III studies[15]. To assess whether the population in the studies and the wider Dutch population are similar, we compared the race distribution in the OAK & POPLAR cohorts with national figures from the Central Bureau of Statistic (CBS) since race is one of the components of ethnicity. Based on such an assessment, health care professionals may decide whether a model is applicable to specific patients. According to the CBS, 23.6% of the population in the Netherlands has a migration background[16]. For this assessment, we assume that all residents with a migration background are non-Caucasian although in fact this is not the case. Furthermore, we assume that everybody in the Netherlands who has a non-migration background is Caucasian.

This assumption is also too generic because there are non-Caucasian people in the Netherlands living here for several generations. These assumptions are used for illustrative purposes based on census data available at this time. To estimate if our test population is representative of the Dutch population, we calculated the percentage of non-Caucasian patients in the cohort. For OAK and POPLAR this is 28.7%, indicating the clinical cohorts have a similar distribution to the wider population. To identify potential bias in our model results, we tested whether the OS model predictions are biased towards the Caucasian race, meaning that the model would perform better when predicting outcomes for Caucasian patients. Therefore, we analyzed how often the OS model makes correct predictions for Caucasian vs. non-Caucasian patients and found that model accuracy for non-Caucasian is 78,5% and accuracy for Caucasian is 71%. Using a Pearson Chi-square Test, we found that this difference is not statistically significant ($p = 0.26$), therefore based on available data the model does not perform differently for different races and thus may have good applicability to cohorts with diverse racial backgrounds.

5 Acknowledgements

We would like to thank the PHC Alliance members Ewoudt van de Garde, Michel van den Heuvel and Daan Hurkmans for sharing their expertise to provide input, review our work, and challenge us. We would also like to thank supporting PHC Alliance members from Mobiquity: Ivo Evers & Teun Schutte, and from PwC: Wouter Karman & Suzanne van Veen for their active participation in the project workshops.

This study was commissioned by Roche Global Product Development Medical Affairs - Personalized Healthcare.

ImmuunPRO Hackathon

Appendices

NB: in addition to the appendices included in this report, scripts from the project code repository may be available upon request.

Appendix A: Data Access Journey

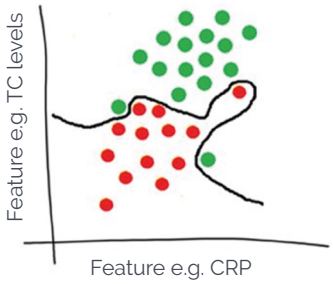
Steps	Timeline	Notes
	may-18 jun-18 jul-18 aug-18 sep-18 oct-18 nov-18 dec-18 jan-19 feb-19 mrt-19 apr-19 mei-19 jun-19 jul-19 aug-19 sep-19 okt-19 nov-19 dec-19 jan-20 feb-20 mrt-20 apr-20 Q4 2020	
1. Proposal		
Initial plan Data Hackathon to generate predictive model for CIT outcomes NSCLC patients		Based on Roche data, Flatiron data, hospital data using Machine Learning algorithm.
Proposal approved and funded by Global PDMA PHC team		
2. Journey		
Preferred 3rd party for preparation & setting up data hackathon event: Accenture		Advised by Roche Global PDMA PHC.
Contracts/ Statement of Work with Accenture		Time consuming process (Oct 2018 - first draft available).
Discussion about NIS		Unexpected discussion. Use of 2nd data initiated by Region Europe Disease Area Medical Lead PHC2.0 &CIT Pharmaceuticals Division took 2 months. Boomerang discussions without a clear conclusion. Finally overruled by local medical director.
3. Data		
Reach out to internal data sources		
Access to Roche data initially a no go from RWD Collaborations PHC Data Science		
Flatiron data a no go from Academic Partnerships/ Network operations and Life Sciences		
List with important atezolizumab studies which potentially could serve the data analysis		Global PDMA Data Sharing Team.
For logistical purpose focus on 2 clinical studies: OAK & POPLAR		Originally 9 studies were used for internal RAAD data challenge Feb 2019.
Biometric MGT & SWAT team objections regarding data access (IT/IP/3rd party competencies)		MGT & SWAT team - Program Management PDP Business Operations Product Development, Personalized Healthcare.
Direct contact with Tecentriq Lung Lifecycle Team Meeting Genetech		
Aug 20th 2019: approval for use of OAK and POPLAR limited datasets		
Clarification which OAK/POPLAR data can be used for analysis		
Data available - PDMA will assist in transfer data to IT environment accessible to Accenture		
4. Contracts/ IP/ IT settings		
Plan for data hackathon, i.e. data mining with combined data is a no go		Reasons include lack of control over model outcomes.
Solution: virtual dataroom		
Decision on legal requirements that need to be in place		Decision that Accenture would access data sets. Data is already anonymized so tasks to be described in SoW. Researchers analyzing data sets in dataroom: DVO type of contract.
IT settings between Roche Global and Accenture are in place		
5. Data analysis in dataroom		
Data preparation & clean by Accenture		
Feature engineering, model build, fine tuning		
Project meeting with Roche, Accenture and Alliance members		
Report on final model		
Report on data analysis outcome		
Action plan to validate algorithm in each hospital (RadboudUMC, ErasmusMC, Santeon)		
First model validation in one of the hospitals		

More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix B: Modelling Approach

This is a supervised machine learning (SML) problem

- Predict PFS at 3 months from baseline
- Predict OS at 2 years from baseline



The output of the SML model is a probability

Information about all the features is combined into the model, which is then used to predict per person the probability of the outcome (OS and PFS, respectively)

Patient ID	Target	Features
1	Yes	24
2	No	41
3	No	47
.	.	.
n	N	n

Patient ID	Target	Features
1	?	24

Patient ID	Probability
1	0.80

Modelling approach

Modelling options can be explored to select the optimal predictive model for the given scope. The options include:

1. Model with features directly correlated with the target
2. Model with clustered features
3. Ensemble model

Different **algorithms** will be tested for each modelling option. These include:

- Random forest
- XGBoost
- GBM

More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix C: Modelling Variables

Nr.	Function to be mapped	Variable	Meaning	Table	Cohort
1	Age	AGE	Age	asl	both
2	Gender	SEX	Sex	asl	both
3	Race	RACE	Race	asl	both
4	Tumor type (squamous vs non-squamous)	HIST	Pathology/Histology (squamous vs. non-squamous)	asl	both
5	Raw PD-L1 scores	TC	TC Raw Score	asl	both
6	Raw PD-L1 scores	IC	IC Raw Score	asl	both
7	PD-L1 scores	TCLEVEL	TC Score (0 to 3)	asl	both
8	PD-L1 scores	ICLVL1	IC Score (0 to 3)	asl	oak
9	PD-L1 scores	IHCLEVEL	IC Score (0 to 3)	asl	poplar
10	ECOG performance status	ECOGGR	ECOG Performance Score	asl	both
11	BMI	BBMI	Baseline Body Mass Index (kg/m2)	asl	both
12	Weight	BWT	Baseline Weight	asl	both
13	Stage	CASTG	Stage of Initial Diagnosis	asl	both
14	Time since diagnosis	XDXDY	Date of Initial NSCLC Diagnosis	asl	both
15	Nr. of prior treatments	PRIORTXC	Prior Therapies per IxRS	asl	oak
16	Nr. of prior treatments	PRIORTX	Prior Therapies per IxRS	asl	poplar
17	EGFR mutations	EGFRMUT	EGFR Mutation Status	asl	both
18	KRAS mutations	KRASMUT	KRAS Mutation Status	asl	both
19	ALK mutations	EMLAMUT	EMLA-ALK Mutation Status	asl	poplar
20	ALK mutations	EML4MUT	EML4-ALK Rearrangement Status	asl	oak
21	Tobacco use history	TOBHX	Tobacco Use History	asl	both
22	Metastasis	METSITES	Number of Metastatic Sites at Enrollment	asl	both
23	Metastasis	BONE	Bone Metastasis at Enrollment	asl	both
24	Metastasis	BRAIN	Brain Metastasis at Enrollment	asl	both
25	Metastasis	PEFF	Pleural Effusion Metastasis at Enroll	asl	both
26	Metastasis	LUNG	Lung Metastasis at Enrollment	asl	both
27	Metastasis	LIVER	Liver Metastasis at Enrollment	asl	both
28	Medical history	MHSEQ	Sequence Number of Medical History Event	amh	both
29	Medical history	MHLLT	Lowest Level Term	amh	both
30	Medical history	MHSOC	Primary System Organ Class	amh	both
31	Medical history	MHSTDY	Study Day of Start of History Event	amh	both
32	Baseline protein levels including albumin	PARAM	Parameter Description	alb	both
33	Lab results	VISITNUM	Visit Number	alb	both
34	Lab results	VISIT	Visit Name	alb	both
35	Comedication dose	CMDOSE	Dose per Administration	acm	both
36	Comedication at baseline	CMCLAS	Medication Class	acm	both
37	Vital signs	PULSE	Pulse	avs	both

The missing value treatment per variable will be included together with the source file (to be included).

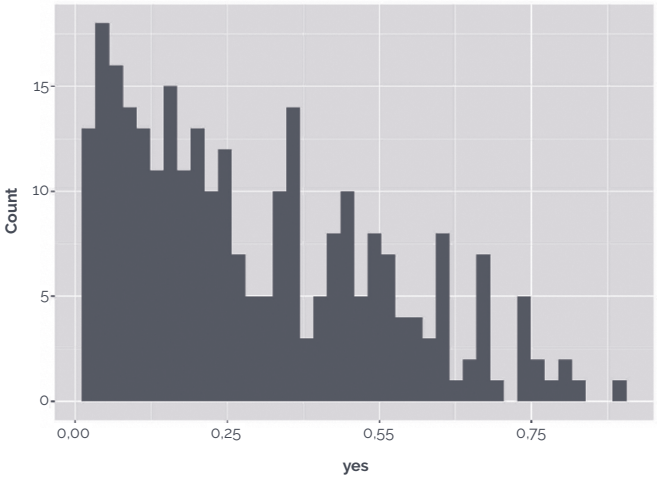
More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix D: OS Model Report

To evaluate the predictive model, we assess the overall performance and the sensitivity and specificity. Sensitivity indicates the percentage of patients correctly predicted by the model to be alive at 2 years and specificity is the percentage of patients correctly predicted not to be alive.

As seen in the modelling approach (Appendix B), the result of a supervised machine learning model is a probability. Since we are predicting an outcome with two possibilities (patient is alive or not), we are making a binary classification. This means that for every new patient where we use the model to make a prediction, the probability returned will be classified as either 0 or 1. When the threshold is set at 0.5, any patient predicted by the model to have probability of survival > 0.5 will be classified as 1, the other patients will be classified as 0. If we change the threshold to 0.4 for example, the patients with probability of survival 0.45 will be classified as 1 and those with probabilities lower than 0.4 will be classified as 0.

Distribution of the probabilities that a patient did survive



At a standard threshold of 0.5, the sensitivity in the confusion matrix is low (0.41) indicating our model underperforms in predicting the less frequent class (the patients that did survive). Normally, a threshold of 0.5 is used when the probability of the outcome is normally distributed between 0 and 1. However, this is not the case in the OAK & POPLAR cohorts since we know most of these patients do not survive. By plotting the histogram of probabilities that a patient survives (Fig.9), we see this distribution is heavily skewed towards the left, with most distributions in the range between 0 and 0.7. By adjusting the threshold to 0.35, both classes of patients - those predicted to be alive and those who are not - are well represented and the model has better sensitivity (0.67) and specificity (0.76). By keeping the threshold at 0.35 for all OS models, different models can be compared against each other.

More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix E: PFS Model Report

More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix F: Antibiotic Analysis

More detailed information is available and can be requested via info@phc-catalyst.nl

Appendix G: Correlation Analysis

The correlation matrices between all variables with non-zero variance with PFS, OS and ordered matrices between targets and all variables is available.

For a visual overview of the matrices above, correlation plots clustered hierarchically are available for each of the data sources, per OS and PFS. The treatment response is not included here because this dataset is only used to define the targets.

ASL – subject level analysis (basic patient characteristics)
ACM – comedication
ALB – lab values
AMH – medical history
AVS – vital signs

More detailed information is available and can be requested via info@phc-catalyst.nl

ImmuunPRO Hackathon

References

References

1. J. S. Sunil Mathur, "Personalized medicine could transform healthcare," *Biomedical Reports*, vol. 7, no. 1, pp. 3-5, 2017.
2. Gupta Strategists, "Personalized Healthcare Alliance - Discussion Paper," 2019. [Online]. Available: https://phc-catalyst.nl/Discussion_Paper/. [Accessed 20 05 2020].
3. John Hopkins Medicine, "Immunotherapy: Precision Medicine in Action," 2020. [Online]. Available: https://www.hopkinsmedicine.org/inhealth/old-template/policy-briefs/immunotherapy-precision-medicine-action-policy-brief.html. [Accessed 20 05 2020].
4. Roche Nederland, "Personalised Healthcare Alliantie," 2019. [Online]. Available: https://www.roche.nl/nl/Over-Roche/Onze-partners/personalised-healthcare-alliantie.html. [Accessed 20 05 2020].
5. Personalized Healthcare Catalyst, "Personalized Healthcare Catalyst - Over Ons," [Online]. Available: https://www.phc-catalyst.nl/#!. [Accessed 20 05 2020].
6. A. Rittmeyer, F. Barlesi, D. Waterkamp, K. Park, F. Ciardiello, J. v. Pawel, S. M. Gadgeel, T. Hida, D. M. Kowalski, M. C. Dols, D. L. Cortinovis, J. Leach and J. Poliko, "Atezolizumab Versus Docetaxel in Patients With Previously Treated Non-Small-Cell Lung Cancer (OAK): A Phase 3, Open-Label, Multicentre Randomised Controlled Trial," *The Lancet*, vol. 389, no. 10066, pp. 255-265, 2017.
7. L. Fehrenbacher, A. Spira, M. Balinger, M. Ballinger, M. Kowanetz, J. Vansteenkiste, J. Mazieres, K. Park, D. Smith, A. Artal-Cortes, C. Lewanski, F. Braiteh, D. Waterkamp and P. He, "Atezolizumab Versus Docetaxel for Patients With Previously Treated Non-Small-Cell Lung Cancer (POPLAR): A Multicentre, Open-Label, Phase 2 Randomised Controlled Trial," *The Lancet*, vol. 387, no. 10030, pp. 1837-1846, 2016.
8. L. Claret, J. Y. Jin, C. Ferte, H. Winter, S. Girish, M. Stroh, P. He, M. Ballinger, A. Sandler, A. Joshi, A. Rittmeyer, D. Gandara, J.-C. Soria and B. Rene, "A Model of Overall Survival Predicts Treatment Outcomes with Atezolizumab versus Chemotherapy in Non–Small Cell Lung Cancer Based on Early Tumor Kinetics," *Clinical Cancer Research*, vol. 24, no. 14, pp. 3292-3298, 2018.
9. A. M. Hopkins, G. Kichenadasse, E. Garrett-Mayer, C. S. Karapetis, A. Rowland and M. J. Sorich, "Development and validation of a prognostic model for patients with advanced lung cancer treated with the immune checkpoint inhibitor atezolizumab," *Clinical Cancer Research*, 2020.
10. M. M. Oken, R. H. Creech, D. C. Tormey, J. Horton, T. E. Davis, E. T. McFadden and P. P. Carbone, "Toxicity and Response Criteria of the Eastern Cooperative Oncology Group," *American Journal of Clinical Oncology*, vol. 5, no. 6, pp. 649-655, 1982.
11. H.-C. Chen, R. L. Kodell, K. F. Cheng and J. J. Chen, "Assessment of Performance of Survival Prediction Models for Cancer Prognosis," *BMC Medical Research Methodology*, vol. 12, no. 102, 2012.
12. P. C. Austin and E. W. Steyerberg, "Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable," *BMC Medical Research Methodology*, vol. 12, no. 82, 2012.
13. M. Chalabi, A. Cardona, D. R. Nagarkar, A. D. Scala, D. R. Gandara, A. Rittmeyer, M. L. Albert, T. Powles, M. Kok and F. G. Herrera, "Efficacy of Chemotherapy and Atezolizumab in Patients With Non-Small-Cell Lung Cancer Receiving Antibiotics and Proton Pump Inhibitors: Pooled Post Hoc Analyses of the OAK and POPLAR Trials," *Annals of Oncology*, vol. 31, no. 4, pp. 525-531, 2020.
14. European Medicines Agency, "ICH Topic E 5 (R1) Ethnic Factors in the Acceptability of Foreign Clinical Data - Step 5 NOTE FOR GUIDANCE ON ETHNIC FACTORS IN THE ACCEPTABILITY OF FOREIGN CLINICAL DATA (CPMP/ICH/289/95)," 2006. [Online]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-5-r1-ethnic-factors-acceptability-foreign-clinical-data-step-5_en.pdf. [Accessed 20 05 2020].
15. P. Shenoy, "Multi-regional clinical trials and global drug development," *Perspectives in Clinical Research*, pp. 62-67, 2016.
16. Centraal Bureau voor de Statistiek, "CBS Open data StatLine," 2020. [Online]. Available: https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS&tableId=37296ned&_theme=61. [Accessed 20 04 2020].