ORACLE

Oracle Health Sciences. For life.

Study Startup Around the World

A quarterly analytical discussion

ChromoReport

SPRING 2020

Clinical operations staff need to have confidence in machine learning predictive models and be able to validate the accuracy of outcomes. By knowing which indicators have the most impact on these models, organizations can focus on those indicators to refine their models and learn from these insights, which can ultimately drive behavioral changes (i.e., less reliance on subjective decisions) to optimize business processes.

Machine learning allows organizations to continuously improve with direct implications on timelines and associated costs of clinical trials. The SHapley Additive exPlanations (SHAP) diagrams allow clinical operation teams to ascertain the importance of indicators, their relative weighing, and interaction. By conducting this analysis, insights and prediction with confidence is possible.

What leading indicators are important and how do you use them in predictive analytics?

To take advantage of predictive analytics, you must first identify what leading indicators have the greatest impact on the process you wish to improve. This has been a challenge in the industry, due to a lack of industry understanding on how machine learning capabilities can be applied to operational metrics.

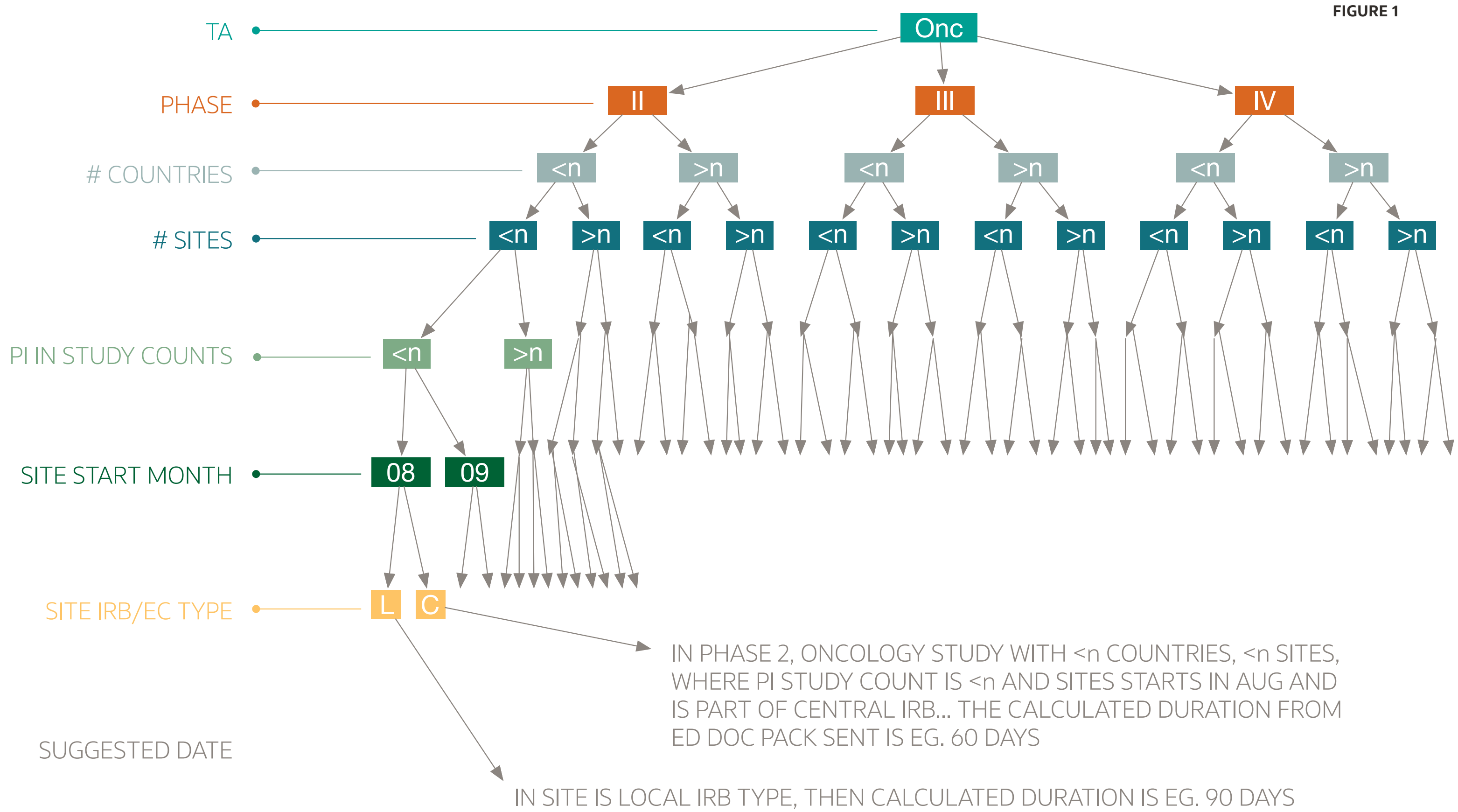
To address this, Oracle has proactively collaborated with sponsors, CROs and industry groups such as the Metrics Champion Consortium (MCC), the TMF Reference Model and others to define a common and agreed-upon list of the most important leading indicators in site activation.

As a result of this collaboration, the leading indicators identified as critical to predicting activation milestones are:

- Therapeutic area
- Phase
- Region code
- Country code
- Countries in study
- Sites in study
- Sites in country
- Start month
- PI's numbers of studies
- IRB/EC type

With leading indicators identified, the machine learning models can then create a decision tree(s) to return the predicted site activation milestone date for a study (Fig 1).

FIGURE 1



As you can see a decision tree(s) allows you to calculate a duration period.

As the indicators in the decision tree could be implemented in a different sequence it is important that the sequence does not impact the weight of each indicator,[1] an important consideration in ensuring confidence and validation of results.

What indicators have the biggest impact on predictions and how do they interact?

SHAP provides a way to graphically reverse-engineer the output of any predictive algorithm, allowing you to understand what decisions the model is making. SHAP represents an importance measure for each indicator,[2] showing their respective impact on the prediction. To illustrate this, Figure 2 depicts the importance of leading indicators for study sites, using a data set of approximately 40,000 sites, to create a summary plot of site activation (IP Release).

In this plot, each study site is represented as a single dot for each indicator under investigation. The horizontal position of the dot is the impact of that indicator on the model's prediction for the study site, and the color of the dot (e.g., red for local IRB and blue for central IRB) represents the value of that indicator for the study site. A negative SHAP value for cycle times is desirable, because that represents a *decrease* in cycle time.

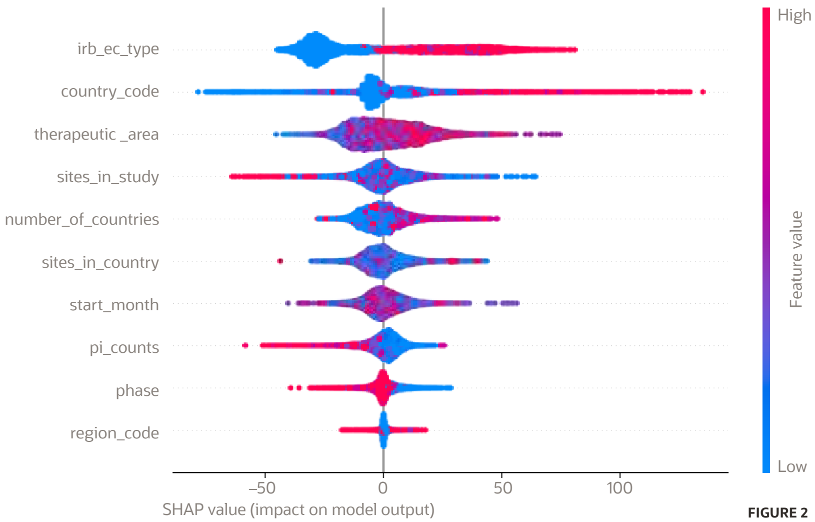


FIGURE 2

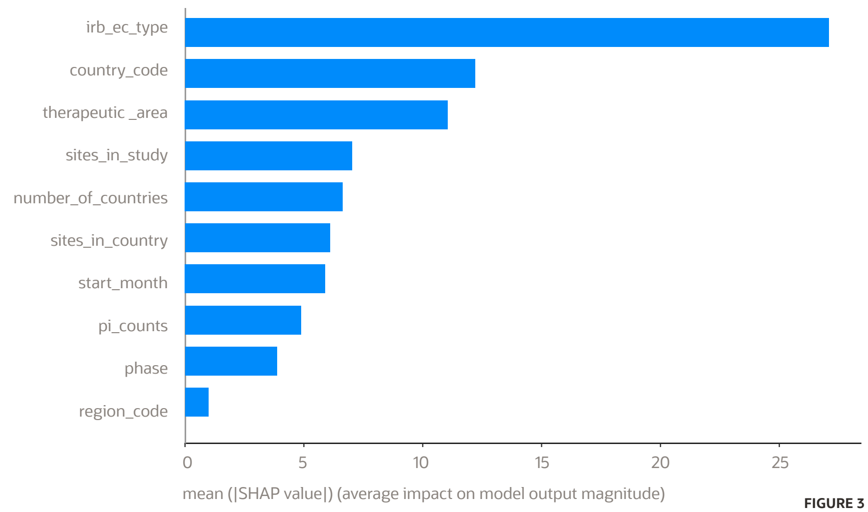


FIGURE 3

As you can see in Figure 2, the highest value indicators in predication of site activation cycle times are the IRB/EC type and country, meaning that these indicators will have the most influence on predicting site activation date. However, to get a more meaningful measure of impact, a simple summary bar plot (Fig 3) shows that IRB/EC type is about two times more important to prediction than the next indicators, which are country and therapeutic area.

In Figure 4, we focus on US sites only, to understand the impact of the IRB/EC type. With the IRB/EC type being the most important feature, there is now a clear separation of central IRB/EC type, showing it is clearly preferable when looking at activation timelines as opposed to local IRB/EC type. The dot clustering for central IRB/EC types in the SHAP plot can be attributed to the consistency in operations of central IRB/ECs.

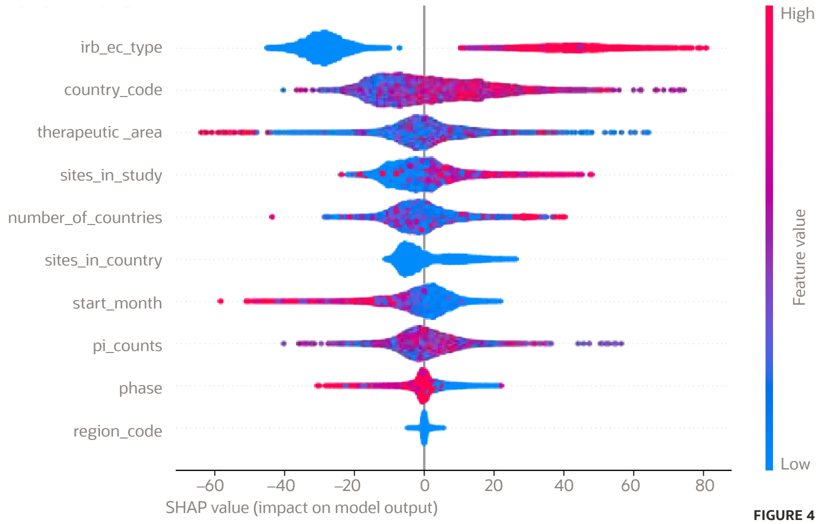


FIGURE 4

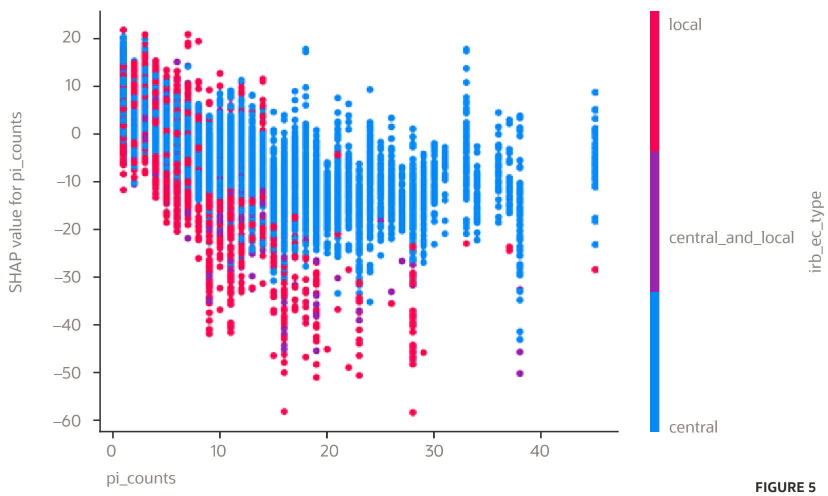


FIGURE 5

As you can see, the two indicators that are directly related to the study site are IRB/EC type and PI count. PI count is a proxy for experience, representing the number of studies that the PI has participated in. With this insight, we can now see how these indicators are interdependent using a dependency plot (Fig 5).

This plot shows a negative SHAP value for experienced primary investigators (study count over 10) using local IRB/ECs, but does that mean that these investigators also activate faster?

The answer is “no.” As the summary bar plot illustrates (Fig 3), the impact of the mean SHAP value for IRB/EC is over five-and-a-half times higher than PI counts, so less experienced investigators using central IRB/EC will, in most cases, activate faster than more experienced investigators using local IRB/EC.

From this example, you can see the value of being able to explore the interactions of leading indicators on predicting site activation.

In Summary

The use of machine learning predictive models[3] is gaining traction in the life science industry to improve key elements in clinical trials. Success, however, relies on finding and refining the right indicators to improve accuracy in outcomes. Machine learning provides critical operational insights, allowing organizations to learn and adapt. Ultimately, these insights allow organizations to transition away from subjective decisions to data-driven decisions, by leveraging these insights to optimize activities in the planning and execution of clinical trials.

Need help?

Oracle Health Sciences, in collaboration with Oracle Labs,[4] is uniquely positioned to assist with the implementation of machine learning capabilities.

References

1 Lundberg, S. Interpretable Machine Learning with XGBoost April 17, 2018 https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27

2 Molnar, C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. June 29, 2020 https://christophm.github.io/interpretable-ml-book/

3 Press Release: Adoption of Artificial Intelligence is High Across Pharmaceutical Industry, According to Tufts Center for the Study of Drug Development, May 7, 2019 https://csdd.tufts.edu/csddnews

4 Oracle Labs. Learn more: https://labs.oracle.com

About Oracle Health Sciences

As a leader in Life Sciences cloud technology, Oracle Health Sciences' Clinical One and Safety One are trusted globally by professionals in both large and emerging companies engaged in clinical research and pharmacovigilance. With over 20 years' experience, Oracle Health Sciences is committed to supporting clinical development, delivering innovation to accelerate advancements, and empowering the Life Sciences industry to improve patient outcomes. *Oracle Health Sciences. For life.*

CONTACT

+1 800 633 0643

healthsciences_ww_grp@oracle.com

www.oracle.com/healthsciences

CONNECT

blogs.oracle.com/health-sciences

facebook.com/oraclehealthsciences

twitter.com/oraclehealthsci

linkedin.com/showcase/oracle-health-sciences