Received: 30 November 2021 | Revised: 30 November 2021 | Accepted: 6 January 2022

DOI: 10.1002/psp4.12760

COMMENTARY

Update to improve reproducibility and interpretability: A response to "Machine Learning for Tumor Growth Inhibition"

Phyllis Chan[1] | **James Lu**[1] | **René Bruno**[2] | **Jin Y. Jin**[1]

[1]Department of Clinical Pharmacology, Roche/Genentech, South San Francisco, California, USA

[2]Department of Clinical Pharmacology, Roche/Genentech, Marseille, France

Correspondence
Phyllis Chan, Department of Clinical Pharmacology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA.
Email: chan.hui-min@gene.com

Funding information
No funding was received for this work

Reproducibility is an important aspect of pharmacometric research, but complications can occur when complex data management and an understanding of modeling approaches from multiple disciplines are required, such as the case when machine learning (ML), a tumor growth inhibition (TGI) model, and baseline covariates were used to predict overall survival (OS) outcome. In addition, continuous updates of our original work,[1] based on more sophisticated ML methodologies, are expected to increase the interpretability of the modeling results.

In the commentary by Meid et al.,[2] the authors originally set out to reproduce our previously published work that explored four ML methods to support the validity of TGI metrics in predicting OS using data from a phase III clinical trial.[1] After eventually obtaining access to the clinical trial data through a data-sharing platform (www.vivli.org), Meid et al.[2] developed an alternative model to predict OS based on the conditional average treatment effect of each patient. To increase the interpretability of the contribution of the covariates in predicting the outcome, Meid et al. also investigated the following two additional methods: variable importance and partial dependence plots.[2]

During our various prior communications with the primary author of Meid et al.[2] to facilitate data access and analysis data set construction, we regret the hurdles that the authors experienced in submitting a research proposal to Vivli (as a simple publication reproduction was not a valid objective in the data request) and constructing an analysis data set from Study Data Tabulation Model (STDM) databases. Dataset building for pharmacometric analyses is not a straightforward process,[3] and the attempt by Meid et al.[2] was also complicated by the lack of readily available TGI metrics in the STDM database, as the individual TGI metrics were model-derived parameters based on modeling of longitudinal tumor size data. In addition, an understanding of TGI models is essential to reproduce the TGI metrics. As alluded to in our publication, a biexponential TGI model[4] was used to fit tumor size data using a nonlinear mixed effect modeling approach.[5,6] An example model file as implemented in NONMEM is available as supporting information in a separate publication on applying the TGI-OS framework to several tumor types[7] and could be helpful for deriving TGI metrics from longitudinal tumor size data in general.

The commentary by Meid et al. also underscored the importance of ensuring ML model transparency and interpretability to clinicians.[2] Our previously published work used Brier scores to compare the predictive performance among models, and the marginal effect of TGI metrics on OS were shown using a simple two-dimensional plot.[1] As mentioned in our publication, to our knowledge, this work that was originally conducted in 2018 was the first work

that leveraged ML to characterize the relationship between TGI metrics and OS. As such, the analysis was intended to be exploratory in nature, based on four well-established ML methodologies (i.e., lasso, boosting, random forest, and kernel machine), and TGI metrics were shown to be the top predictors of OS, regardless of the method used.

We agree with Meid et al.[2] regarding the importance of ML model interpretability and also considered alternative, more sophisticated model explanatory methodologies to show the impact of covariates on the outcome. As a result, an extended analysis has been conducted, and the results were presented at the annual meeting of American Society for Clinical Pharmacology and Therapeutics in March 2021.[8] In this subsequent work, an XGBoost model identified a parsimonious feature set, which also included TGI metrics among the top predictors, and SHapley Additive exPlanation (SHAP) values[9] enabled visualization of the marginal effects from the covariates on OS in a more comprehensive fashion. As shown in Sundrani and Lu,[10] SHAP analysis delivers not only a ranked order of variable importance but also provides a detailed attribution of variable contribution to the predicted hazard rates at the individual patient level. As a result, the SHAP methodology has become widely adopted for biomedical applications to enable better interpretability of ML models. A full manuscript is planned to follow with details on the methods of the updated analysis and additional results such as predicted OS hazard ratios to reflect the relative treatment benefit as well as their confidence intervals that can be computed from ML models,[10] which would extend further the interpretability and practical applicability of our work.

The increasing popularity and accessibility of ML approaches in recent years has prompted us to investigate their applicability on leveraging TGI metrics and baseline covariates to predict OS outcome. With the continuous improvement and development of new ML algorithms and the availability of large data sets, additional iterations and updates of our original work are expected in the future.

CONFLICT OF INTEREST

The authors are employees and stockholders of Genentech, Inc.

REFERENCES

1. Chan P, Zhou X, Wang N, et al. Application of Machine Learning for Tumor Growth Inhibition—overall survival modeling platform. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:59-66.
2. Meid AD, Gerharz A, Groll A. Machine Learning for Tumor Growth Inhibition–interpretable predictive models for transparency and reproducibility. *CPT Pharmcometrics Syst Pharmacol*. In press.
3. Radivojevic A, Corrigan B, Downie N, et al. Data standards for model-informed drug development: an ISoP initiative. *J Pharmacokinet Pharmacodyn*. 2018;45:659-661.
4. Stein WD, Gulley JL, Schlom J, et al. Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. *Clin Cancer Res*. 2011;17:907-917.
5. Claret L, Gupta M, Han K, et al. Evaluation of tumor-size response metrics to predict overall survival in Western and Chinese patients with first-line metastatic colorectal cancer. *J Clin Oncol*. 2013;31:2110-2114.
6. Claret L, Jin JY, Ferté C, et al. A model of overall survival predicts treatment outcomes with atezolizumab versus chemotherapy in non-small cell lung cancer based on early tumor kinetics. *Clin Cancer Res*. 2018;24:3292-3298.
7. Chan P, Marchand M, Yoshida K, et al. Prediction of overall survival in patients across solid tumors following atezolizumab treatments: a tumor growth inhibition-overall survival modeling framework. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1171-1182.
8. Duda M, Chan P, Bruno R, Jin JY, Lu J. A pan-indication machine learning (ML) model for tumor growth inhibition-overall survival (TGI-OS) prediction. *Clin Pharmacol Ther*. 2021;109:S25.
9. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67.
10. Sundrani S, Lu J. Computing the hazard ratios associated with explanatory variables using machine learning models of survival data. *JCO Clin Cancer Inform*. 2021;5:364-378.