

## ARTICLE

# Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN Gastric 100 phase III trial of avelumab

Nadia Terranova<sup>1</sup> | Jonathan French<sup>2</sup> | Haiqing Dai<sup>3</sup> | Matthew Wiens<sup>2</sup> | Akash Khandelwal<sup>4</sup> | Ana Ruiz-Garcia<sup>5</sup> | Juliane Manitz<sup>3</sup> | Anja von Heydebreck<sup>4</sup> | Mary Ruisi<sup>3</sup> | Kevin Chin<sup>3</sup> | Pascal Girard<sup>1</sup> | Karthik Venkatakrishnan<sup>3</sup>

<sup>1</sup>Merck Institute of Pharmacometrics (an affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland

<sup>2</sup>Metrum Research Group, Tariffville, Connecticut, USA

<sup>3</sup>EMD Serono, Billerica, Massachusetts, USA

<sup>4</sup>The healthcare business of Merck KGaA, Darmstadt, Germany

<sup>5</sup>Metrum Research Group, San Diego, California, USA

## Correspondence

Nadia Terranova, Merck Institute for Pharmacometrics, EPFL Innovation Park – Building I, CH-1015 Lausanne, Switzerland.  
Email: nadia.terranova@merckgroup.com

## Funding information

Funding for the JAVELIN Gastric 100 trial and analyses reported here was provided by the healthcare business of Merck KGaA, Darmstadt, Germany (CrossRef Funder ID: 10.13039/100009945), as part of an alliance between the healthcare business of Merck KGaA, Darmstadt, Germany, and Pfizer. Medical writing support was provided by Felicia Barklund of Clinical Thinking and funded by the healthcare business of Merck KGaA, Darmstadt, Germany, and Pfizer

## Abstract

Avelumab (anti-PD-L1) is an approved anticancer treatment for several indications. The JAVELIN Gastric 100 phase III trial did not meet its primary objective of demonstrating superior overall survival (OS) with avelumab maintenance versus continued chemotherapy in patients with advanced gastric cancer/gastroesophageal junction cancer; however, the OS rate was numerically higher with avelumab at timepoints after 12 months. Machine learning (random forests, SIDEScreen, and variable-importance assessments) was used to build models to identify prognostic/predictive factors associated with long-term OS and tumor growth dynamics (TGDs). Baseline, re-baseline, and longitudinal variables were evaluated as covariates in a parametric time-to-event model for OS and Gompertzian population model for TGD. The final OS model incorporated a treatment effect on the log-logistic shape parameter but did not identify a treatment effect on OS or TGD. Variables identified as prognostic for longer OS included older age; higher gamma-glutamyl transferase (GGT) or albumin; absence of peritoneal carcinomatosis; lower neutrophil-lymphocyte ratio, lactate dehydrogenase, or C-reactive protein (CRP); response to induction chemotherapy; and Eastern Cooperative Oncology Group performance status of 0. Among baseline and time-varying covariates, the largest effects were found for GGT and CRP, respectively. Liver metastasis at re-baseline predicted higher tumor growth. Tumor size after induction chemotherapy was associated with number of metastatic sites and stable disease (vs. response). Asian region did not impact OS or TGD. Overall, an innovative workflow supporting pharmacometric modeling of OS and TGD was established. Consistent with the primary trial analysis, no treatment effect was identified. However, potential prognostic factors were identified.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The healthcare business of Merck KGaA, Darmstadt, Germany and Metrum Research Group. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics

### Study Highlights

#### WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Immune checkpoint inhibitors have limited antitumor activity in patients with gastric cancer or gastroesophageal junction cancer (GC/GEJC), but a subset of patients obtain clinical benefits. Pharmacometric modeling can identify patient and disease factors associated with outcome, potentially enabling more efficient selection of therapy.

#### WHAT QUESTION DID THIS STUDY ADDRESS?

This analysis aimed to identify prognostic or predictive factors in patients with advanced GC/GEJC who received maintenance therapy with avelumab or continued chemotherapy in the JAVELIN Gastric 100 trial. Machine learning was used to develop improved models of overall survival (OS) and tumor growth dynamics (TGD).

#### WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

Advanced models of OS and TGD were generated, and potentially important baseline and longitudinal covariates that are prognostic for OS and TGD were identified.

#### HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

This study presents an innovative workflow for modeling OS and TGD in the presence of covariate sets during immune checkpoint inhibitor therapy, which may be applicable to disease settings beyond GC/GEJC.

## INTRODUCTION

Avelumab, an anti-PD-L1 immune checkpoint inhibitor, has been approved in various countries as monotherapy for metastatic Merkel cell carcinoma and advanced urothelial carcinoma (first-line maintenance and second-line therapy) and in combination with axitinib for advanced renal cell carcinoma.<sup>1-4</sup> JAVELIN Gastric 100 was a phase III trial that compared maintenance avelumab treatment versus continuation of chemotherapy in patients with advanced gastric cancer or gastroesophageal junction cancer (GC/GEJC) that had not progressed with first-line induction chemotherapy (Figure 1).<sup>5</sup> This trial did not meet its primary objective of demonstrating superior overall survival (OS) in the avelumab arm. However, in Kaplan-Meier analyses, OS curves crossed after ~ 12 months, with numerically higher OS rates observed in the chemotherapy arm before 12 months and in the avelumab arm after 12 months.<sup>5</sup> These results raised questions about whether a subpopulation exists that could benefit from avelumab.

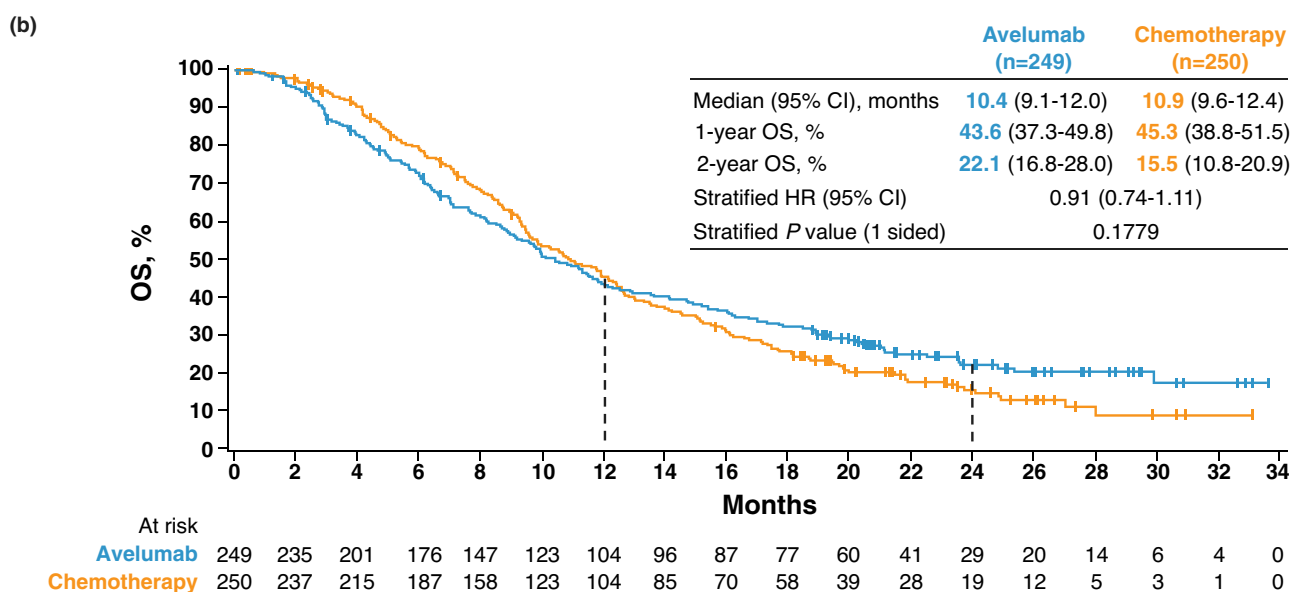
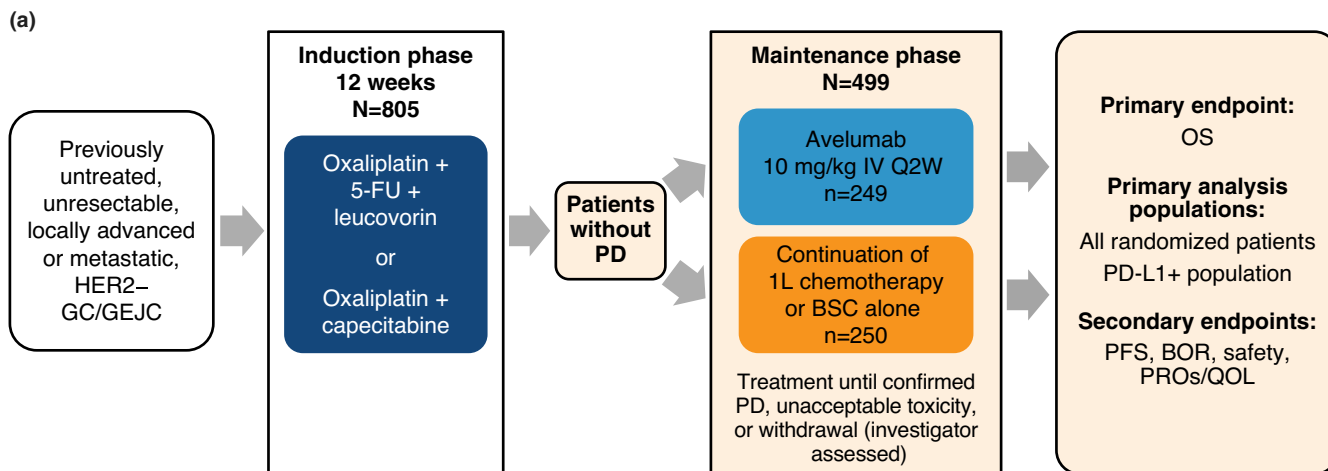
Machine learning (ML) methods are commonly used to identify potential prognostic and predictive factors from trial data, with increasing application in the context of pharmacometric covariate modeling.<sup>6-10</sup> The current analysis aimed to develop pharmacometric models that applied ML approaches to JAVELIN Gastric 100 trial data, with the purpose of identifying prognostic or predictive

factors for OS and tumor growth dynamics (TGDs; i.e., changes in tumor size over time).

## MATERIALS AND METHODS

### JAVELIN Gastric 100

The design and outcomes of the JAVELIN Gastric 100 trial have been reported previously.<sup>5</sup> In brief, the study enrolled 805 patients with unresectable, human epidermal growth factor receptor 2–negative (*ERBB2*–negative), locally advanced or metastatic GC or GEJC who received 12 weeks of first-line induction chemotherapy with oxaliplatin + a fluoropyrimidine. Subsequently, 499 patients without progressive disease (complete or partial response or stable disease) were randomized 1:1 to receive avelumab 10 mg/kg every 2 weeks ( $n = 249$ ) or continued chemotherapy ( $n = 250$ ), stratified by region (Asia vs. non-Asia). All patients received best supportive care (BSC); patients in the chemotherapy arm considered ineligible for further chemotherapy received BSC only ( $n = 19$ ). The primary end point was OS (defined as time from randomization to death). When referring to study data, “baseline” refers to values obtained before the 12-week induction period and “re-baseline” refers to values obtained at randomization (after induction).



**FIGURE 1** JAVELIN Gastric 100 trial overview. (a) Study design schematic. (b) Primary analysis of OS.<sup>5</sup> Borrowed with permission from Moehler M, et al. Phase III trial of avelumab maintenance after first-line induction chemotherapy versus continuation of chemotherapy in patients with gastric cancers: results from JAVELIN Gastric 100. *J Clin Oncol*. 2021;39(9):966–977. <https://ascopubs.org/doi/full/10.1200/JCO.20.00892> © American Society of Clinical Oncology. 1L, first line; 5-FU, 5-fluorouracil; BOR, best overall response; BSC, best supportive care; CI, confidence interval; GC/GEJC, gastric/gastroesophageal junction cancer; HER2, human epidermal growth factor receptor 2; HR, hazard ratio; IV, intravenous; OS, overall survival; PD, progressive disease; PFS, progression-free survival; PRO, patient-reported outcome; Q2W, every 2 weeks; QOL, quality of life; R, randomization; RECIST, Response Evaluation Criteria in Solid Tumors

## Analysis overview

Analyses focused on efficacy outcomes (OS and TGD) for all randomized patients. Patient and disease characteristics at baseline or re-baseline were evaluated as potential predictors. Longitudinal values of laboratory biomarkers (i.e., those assessed over time) were considered for inclusion in the ML models. For time-invariant values, re-baseline values were included as covariates. For time-varying values, change from baseline to re-baseline and

individual-predicted post-randomization annual rates of change (model described later) were included as covariates in the ML model, whereas actual profiles of those identified as relevant were considered in subsequent parametric modeling. In this analysis, we screened 32 time-invariant covariates and 19 time-varying covariates. This yielded a total of 89 and 52 covariates entering the ML models for OS and TGD, respectively. Modeling methodology and software are described in the Supplementary Methods.

## Stability assessment of covariate candidates

Both time-invariant and longitudinal covariates were included in the ML models. To assess longitudinal covariate stability over time, a linear mixed effect model was fitted to post-randomization values (Supplementary Methods). A covariate was considered time invariant if 50% or more patients in each treatment arm had no meaningful trend over time (defined as an individual-specific annual rate of change larger than the estimated residual error standard deviation). For time-invariant biomarkers, re-baseline values were used as covariates in subsequent models. For time-varying variables, baseline, change from baseline to re-baseline, and individual-specific annual rate of change were used as covariates (features) in subsequent ML models. Time-invariant demographic covariates used baseline values.

## Identification of prognostic and predictive markers for OS

To evaluate the relationship between candidate covariates and OS, exploratory graphical analysis was performed. Prognostic markers were identified using random forests (RFs) with tuned parameters.<sup>11</sup> Missing data were imputed to the median for continuous covariates and the mode for discrete covariates, stratified by treatment arm. No imputation was performed for time-varying covariates.

Predictive markers were identified using methods appropriate for exploratory subgroup discovery.<sup>12</sup> In particular, the SIDEScreen procedure (a two-stage algorithm for identifying combinations of predictors, i.e., regions in the covariate space<sup>13</sup>) was used to identify subpopulations showing significant treatment effects.

For prognostic models, covariate importance was assessed using a combination of the Boruta method,<sup>14</sup> the permutation method, and a random splits method. The assessment criteria aligned with the overall hypothesis-generating strategy; therefore, the general approach was to include covariates that one method deemed important, while discarding those that neither method selected. The Boruta method is a strategy in RF models in which randomly permuted copies of the original covariates are introduced as candidate predictors in addition to the original covariates. Because the permuted variables have no relationship to the outcome, the original covariates that are selected more frequently than the permuted covariate are considered to be important predictors. The Boruta method classifies covariates into three categories: confirmed, tentative, and rejected. If not rejected by the Boruta method, a covariate was considered important and

included in the parametric model if the importance determined by either the permutation or random splits method was greater than or equal to 25% of the most important variable identified (relative importance). If the covariate was rejected by the Boruta method, then to be included in the parametric model, the relative importance determined by the permutation and random splits methods needed to be 10% or greater for both methods and 25% or greater for one method. For further evaluations, this process was repeated, excluding the slopes of time-varying covariates to assess variables present only at baseline or re-baseline. For evaluating predictive markers, the SIDEScreen procedure had a built-in measure of covariate importance.

## Development of a parametric time-to-event model

A parametric time-to-event (TTE) model for OS was constructed with re-baseline and time-varying factors. Four structural models for the baseline survival distribution were evaluated: exponential, Weibull, log-logistic, and Gompertz. Time-invariant and re-baseline covariates were evaluated for whether they modified the scale parameter of each model, and time-varying covariates were evaluated for whether they modified the baseline hazard proportionally. Two choices for how covariates entered the model were considered: nonparallel hazard functions to reproduce the crossing of survival curves through covariate effects on the scale parameter, and ensuring a continuous cumulative hazard function by using proportional hazards effects for time-varying covariates.<sup>15</sup> Equations describing such relationships are reported in the Supplementary Methods.

Covariate effects on the shape and scale parameters were assumed to be log-linear. Nonlinear associations were evaluated if model diagnostics indicated that the linear association was not appropriate. A treatment effect on the shape parameter was also explored based on model diagnostic and exploratory plots. Uninformative priors were used in the Bayesian modeling. Credible intervals were derived from percentiles of the posterior distributions of parameters or quantities of interest.

Clinical trial simulations were used to predict landmark survival, accounting for variability in the study population and uncertainty in model parameters. One thousand clinical trials were simulated by resampling the patient population and sampling the parameters from the posterior distribution. The OS simulations accounted for administrative censoring at the end of the trial but not patient discontinuations. Arms of equal size to the original arms were created. Differences in survival probability were computed for landmarks of 12, 18, and 24 months.

## Identification of prognostic and predictive markers for TGD

The fundamental approach for identifying prognostic and predictive markers for TGD was to use ML (RFs and variable importance metrics) approaches on the empirical Bayes estimates (EBEs) of structural parameters in a tumor dynamic model to identify important covariates via three steps. First, a nonlinear mixed-effects model was constructed for longitudinally measured TGD data. Different tumor dynamic models using ordinary differential equations were explored. The base structural model did not include any drug-induced decay process; thus, the net growth characterized by a constant rate of growth accounted for natural tumor growth and any tumor inhibition by therapy. Model evaluations utilized standard procedures in population pharmacology modeling.<sup>16,17</sup>

In the second step, RF methods were used to identify prognostic and predictive factors associated with TGD parameters from a set of 52 covariates. The outcome modeled was individual-predicted (EBE) parameters from the TGD model. An RF model for each outcome was used. Variables were considered prognostic on a random effect if the estimated Shapley importance was greater than 5% of the interquartile range (IQR) of the random effect. The predictive importance of each covariate was assessed using a differential effect of the Shapley values. Each covariate was discretized into quartiles if there were more than four distinct values. Shapley values for each quartile within each arm were averaged to calculate a differential treatment effect of the covariate in each quartile. If the greatest differential treatment effect across the four quartiles was greater than 10% of the IQR, the covariate was considered predictive.

Finally, the base model was updated to include the factors identified in the ML step, resulting in the “final” TGD model. Both scientific understanding of the mechanistic effect of covariates and visual inspection of the relationship between the sampled EBEs and covariates guided selection of parameter-covariate relationships.

## RESULTS

### Identification of prognostic and predictive markers for OS and development of a parametric TTE model

Of 23 longitudinal laboratory values assessed, six were determined to be time invariant and 17 time varying (Table S1). Although large changes were recorded at the last assessment for many patients, these time points had minimal effects on the cumulative hazard in subsequent

modeling. When stratifying by event time quartile, changes in the laboratory values from baseline were much larger in the first quartile than in the fourth quartile, suggesting that changes in laboratory values were associated with OS. Variables concluded to be time-varying (at least one arm containing >50% of patients with more than one standard deviation change per year) were albumin (ALB), alkaline phosphatase (ALP), alanine aminotransferase, aspartate aminotransferase (AST), C-reactive protein (CRP), gamma-glutamyl transferase (GGT), lactate dehydrogenase (LDH), hemoglobin, lymphocytes, lymphocyte-leukocyte ratio, neutrophils, platelets, leukocytes, estimated glomerular filtration rate (eGFR), platelet-lymphocyte ratio, neutrophil-lymphocyte ratio, and inflammation index.

### ML covariate screening

Important covariates identified by ML as prognostic were mostly the estimated slopes from time-varying laboratory values. Among variables observable at re-baseline, Eastern Cooperative Oncology Group performance status (ECOG PS) and age were identified as important (Figure 2a and Table S2). To focus on time-invariant covariates observable before maintenance therapy, the importance of variables available only at re-baseline was also considered. Additional variables were shown to be important; however, the importance of re-baseline ECOG PS and age remained, suggesting that the selection was consistent (Figure 2b). The importance of the identified variables was confirmed by visual inspection of Kaplan-Meier curves for OS, given the stratification of variables into quartiles. The important time-varying covariates included in the parametric model were ALB, ALP, AST, CRP, inflammation index, LDH, lymphocyte-leukocyte ratio, neutrophil-lymphocyte ratio, and platelet-lymphocyte ratio. The important time-invariant covariates included in the parametric model were age, re-baseline ECOG PS, baseline sum of longest diameters (SLDs), creatine kinase (CK), days since diagnosis, eGFR, GGT, heart rate, prior gastrectomy, systolic blood pressure, triglycerides, and tumor response at re-baseline.

Several of the time-varying laboratory value slopes were highly correlated. In particular, absolute correlations between the slopes of platelet-lymphocyte ratio, neutrophil-lymphocyte ratio, lymphocyte-leukocyte ratio, and inflammation index were greater than 0.70. Thus, based on clinical considerations,<sup>18–21</sup> only the neutrophil-lymphocyte ratio was retained for the parametric model.

The SIDEScreen algorithm analysis did not identify any potentially predictive covariates. Initially, simple subgroup analysis identified suggested treatment-modifying



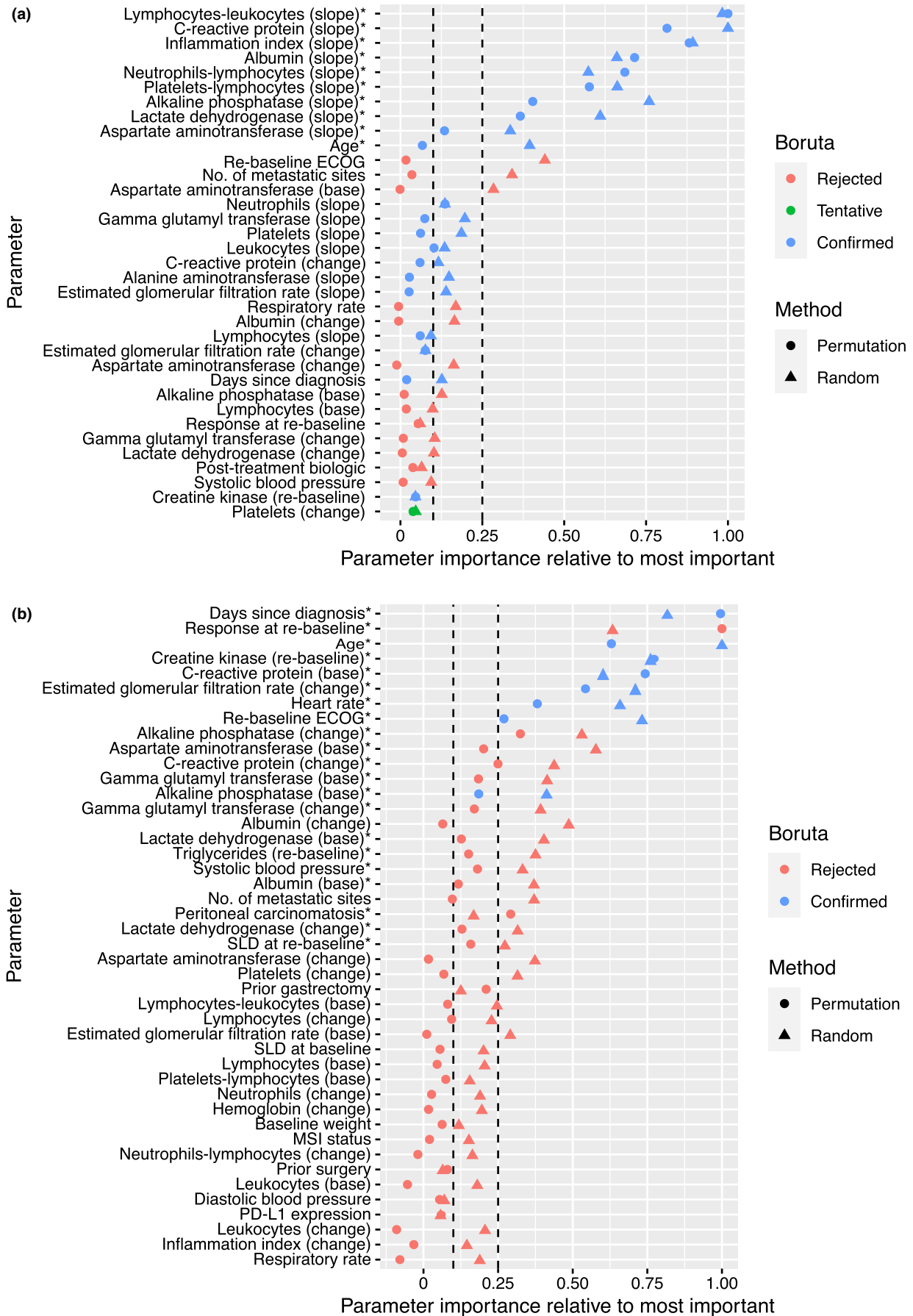


FIGURE 2 Legend on next page

**FIGURE 2** ML variable importance for OS. (a) Relative variable importance per methodology. Covariates included in this plot are filtered to those covariates that the Boruta method does not reject, or those covariates where the mean of the relative importance of the other two methods is greater than or equal to 0.05. (b) Variable importance for those covariates measurable at re-baseline or baseline, relative to the most important covariate for each method. Covariates included in this plot are filtered to those covariates that the Boruta method does not reject, or those covariates where the mean of the relative importance of the other two methods is greater than or equal to 0.05. In both (a) and (b), the vertical bars show the 10% and 25% relative importance thresholds. \*Variables meeting the criteria for consideration in the parametric TTE model. ECOG, Eastern Cooperative Oncology Group; ML, machine learning; MSI, microsatellite instability; OS, overall survival; SLD, sum of the longest diameter of target lesions

effects, but these did not withstand adjustment for multiplicity. As an additional method to assess predictive factors, prognostic factors were assessed separately for each arm; however, identified variables were consistent. Although the order of variable importance among the three data sets (avelumab arm, chemotherapy arm, and both arms combined) differed, no difference in ranking was large enough to suggest a predictive effect.

### Parametric TTE model

The log-logistic model provided the best fit to the OS data. Exploration of the baseline hazard function showed that the log-logistic model best fit the nonmonotonic estimated hazard curve for trial arms, both combined and individually. Maximum likelihood fits for the log-logistic model suggested no treatment effect on median OS. Because the nonparametric kernel hazard estimator, with an Epanechnikov kernel and local optimization of the bandwidth, suggested different shapes of the hazard function for treatment type, a parameter for this effect was also included. ML screening identified both time-invariant (on-scale parameter/accelerated failure time: treatment, age, CK, days since diagnosis, re-baseline ECOG PS, eGFR, heart rate, GGT, peritoneal carcinomatosis, prior gastrectomy, systolic blood pressure, SLD at baseline, triglycerides, and tumor response at re-baseline) and time-varying (on-hazard/proportional hazards using longitudinal values: ALP, AST, CRP, LDH, ALB, and neutrophil-lymphocyte ratio) covariates (Table S3). The visual predictive check (VPC) for the model showed close alignment to observed Kaplan-Meier curves (Figure S1) and hazard ratio (Figure S2). In a substantial fraction of the posterior draws, both the mean hazard function and survival curves exhibited crossing of curves during follow-up, replicating observed OS data. Different shape parameters and covariate distributions contributed to the crossing of the survival curves. Asian versus non-Asian region was not identified as a covariate; VPCs stratified by region are shown in Figure S3.

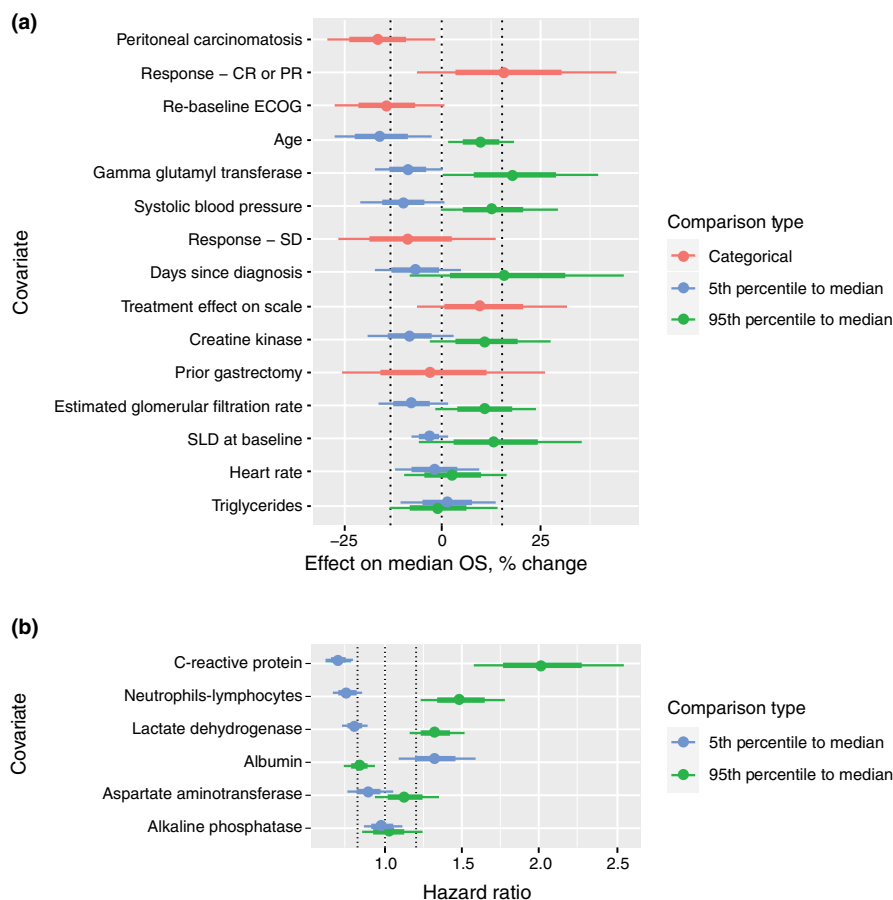
Estimated parameters suggested that the effect of treatment on median OS (i.e., effect on scale parameter) was negligible because the point estimate was close to zero and the credible intervals extended to both positive and

negative effects (Figure 3). Furthermore, the credible interval for the treatment effect on the shape parameter excluded zero; thus, the probability of a non-negligible difference in the shape of the hazard function was high. The magnitude of effects on OS of covariates identified as prognostic via ML were evaluated further. Specifically, the effects of covariates were scaled to the effect on median OS between the 5th and 95th percentiles of continuous covariates, or a reference value of zero or false for categorical covariates. Among time-invariant covariates, the largest relative effect was for GGT, with a median difference of 17% longer OS (1.42 vs. 1.21 years; 95th percentile to median). Among time-varying covariates, the largest effect was for CRP, with median hazard ratios of 0.699 for a CRP of 0.50 versus 2.0 mg/L (5th percentile to median) and 2.01 for a CRP of 34.9 versus 2.0 mg/L (95th percentile to median). Reference and comparison values for the time-varying covariates were derived from observed data at re-baseline.

Time-varying covariates were evaluated using percentiles at re-baseline for comparison. Important variables included neutrophil-lymphocyte ratio, LDH, CRP, albumin, peritoneal carcinomatosis, and age (Figure 3). Model fit was also assessed through stratified VPCs for time-invariant covariates. A possible lack of model fit was seen only in the VPC for heart rate, whereas the others showed the observed data within the 95% credible interval range for predicted OS curves.

When evaluating the effect size and importance for time-varying covariates, trajectories were meaningful, and extreme laboratory values and events were correlated. The effect for CRP and neutrophil-lymphocyte ratio is shown in Figure 4 and Figure S4, respectively. The rapid changes in laboratory values close to event times were also demonstrated by stratifying by first versus fourth quartile of observed event time and assessing median trajectory. The first-quartile (shorter OS) had more rapid changes in laboratory values than the fourth-quartile (longer OS) group. Changes were similar between treatment arms.

The predicted effect of avelumab on long-term OS was evaluated through clinical trial simulation. The probability of longer OS in the avelumab arm increased over time, from 47% at 12 months after randomization to 76% at 24 months. Indeed, estimated differences between the avelumab and chemotherapy arms were predicted to increase



**FIGURE 3** Estimated covariate effects on OS. (a) Estimated covariate effects on median survival for baseline, re-baseline, and stable covariates (dots) and 70% (thick line) and 95% (thin line) credible intervals for each time-invariant covariate effect parameter estimated in the model. Larger estimates correspond to longer survival. For continuous covariates, the 2.5th and 97.5th percentiles are compared with the median. For treatment, the reference is chemotherapy. For the tumor response covariate, the effects are compared with patients who had neither stable disease nor response. For re-baseline ECOG, the effect is compared with a re-baseline ECOG score of 0. For prior gastrectomy, the baseline reference is no prior gastrectomy. For peritoneal carcinomatosis, the reference is no peritoneal carcinomatosis. The dashed lines are at zero, for no effect, and the bounds at  $\pm 15\%$  for the posterior median to classify the variable as having a meaningful effect. (b) Estimated covariate effects on the hazard ratio for time-varying covariates (dots) and 70% (thick line) and 95% (thin line) credible intervals for each time-varying (longitudinal data) covariate effect parameter estimated in the model. Smaller estimates correspond with longer survival. The percentiles for reference and comparison values were calculated using the values of the covariate at re-baseline only. The dashed lines are at one, for no effect, and the bounds  $\pm 1.2$  for the posterior median to classify the variable as having a meaningful effect. CR, complete response; ECOG, Eastern Cooperative Oncology Group; OS, overall survival; PR, partial response; SD, stable disease; SLD, sum of the longest diameter of target lesions

over time (median difference: 12 months,  $-0.40\%$ , 95% confidence interval [CI],  $-0.13$  to  $0.12$ ; 18 months,  $3.2\%$ , 95% CI,  $-0.088$  to  $0.16$ ; and 24 months,  $4.0\%$ , 95% CI,  $-0.060$  to  $0.15$ ); however, all included the null value and were well below the 10% difference considered clinically meaningful (Figure 5 and Figure S5; Table S4).

## Identification of prognostic and predictive markers for TGD

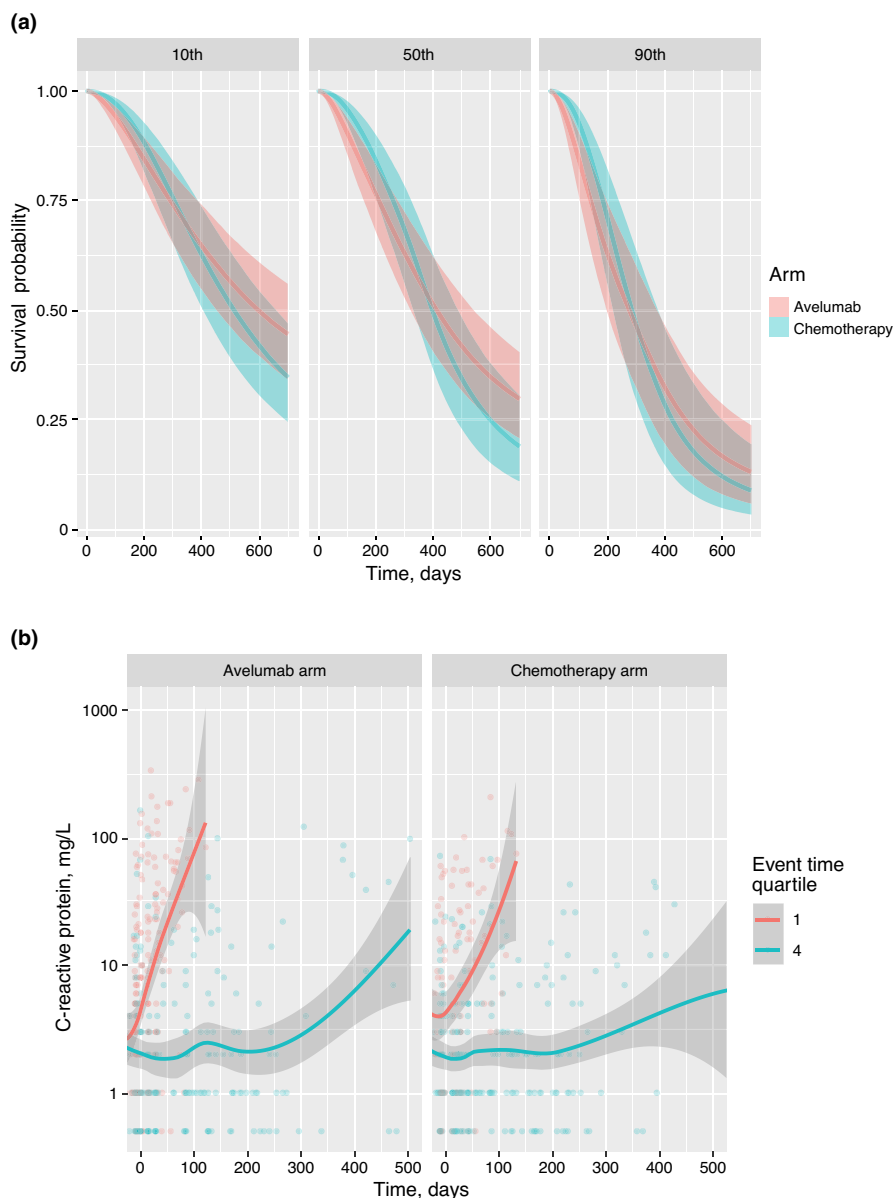
The goodness of fit and stability of different models for TGD were evaluated. The Gompertzian equation was the

most parsimonious model for providing a good description of the data. The base model generally predicted the central tendency in TGD data across patients. During the estimation process, a high correlation was observed between interindividual variability (IIV) on  $K_g$  and  $K_d$  (rate constants of growth and deceleration/decline, respectively; the deceleration rate relates to natural death of tumor cells). Thus,  $K_d$  was reformulated as a parameter proportional to  $K_g$  with a slope and intercept, which allowed the simplification of the dimensionality of the omega structure restricting the correlation between  $K_d$  and  $K_g$  to be 1.<sup>22</sup>

The relationship between the two estimated random effects ( $K_g$ - $K_d$  and baseline tumor size) suggested



**FIGURE 4** Analysis of C-reactive protein (CRP) in relation to overall survival. (a) Estimated survival curves for smoothed 10th percentile, median, and 90th percentile CRP levels over time. All other variables held constant at reference levels. Increases in CRP are estimated to have increased risk. Bands are 95% credible intervals. (b) Observed data for CRP. The first and fourth quartiles of known event times are compared to show the trajectories of CRP for short-term versus long-term survivors



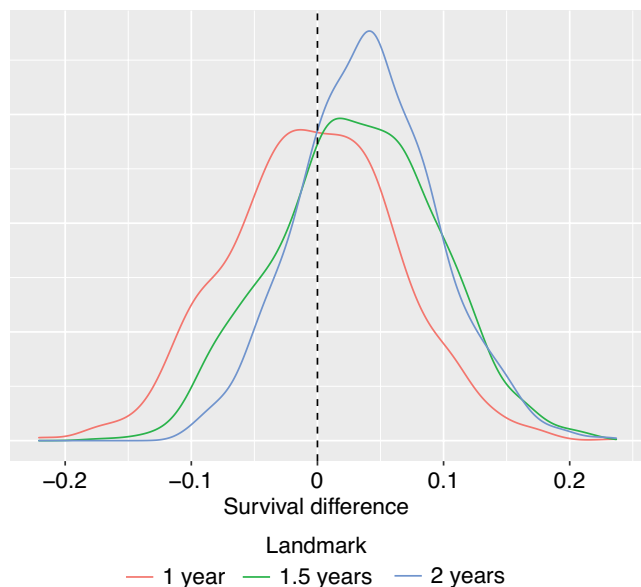
minimal correlation; thus, the two univariate models would not lead to nonidentification of interactions between the random effects. The RF had modest predictive power (19% variance explained) on the baseline TGD random effect and weaker predictive power (4% variance explained) on the growth-deceleration random effect, but this was sufficient to compute variable importance scores. All variables considered in the model were assessed with Shapley values, the permutation algorithm, and the random splits algorithm. The last two methods did not suggest any covariates that were not already captured with Shapley values (Figure S6). No treatment-covariate interactions were included in the final model because each of the differential treatment effects was much less than 10% of the IQR.

Covariate effects on the random effects of the growth-deceleration rate were liver metastasis and time since

diagnosis. For the random effect of baseline TGD, covariates identified were number of metastatic sites and tumor response at re-baseline.

Identified prognostic and predictive markers on TGD dynamics by ML were incorporated to the base model (Table S5) as described in Supplementary Methods to generate the “final” TGD model (Table S6). The objective function value for the “final” TGD model was ~ 121 points lower than for the base TGD model. The effect of liver metastasis is illustrated in Figure S7. The mean percentage change in TGD at the end of the study (Table S7) was calculated by averaging shrinkage (<0%) and growth (>0%). Consistent with the OS model, Asian versus non-Asian region was not identified as a covariate for the TGD model (Figure S8).

Parameter estimates from the final model and median and mean values from the nonparametric bootstrap appeared to agree for Kg and tumor size at baseline and covariate effects



**FIGURE 5** The posterior distributions of survival differences between the avelumab arm and the chemotherapy/best supportive care arm. The probability of the avelumab arm having longer survival increased with time, but there was non-negligible probability of the avelumab arm having shorter survival at all landmark times

on tumor baseline (Table 1). However, estimates for covariate effects on Kg/Kd seemed to present very high uncertainty, with a point estimate that differed between final model estimate and median bootstrap analysis. Stratified VPCs were generated using the final model parameters to determine if observed data were consistent with model-simulated median, 2.5th, and 97.5th percentiles, which indicated acceptable performance of the model in all covariate groups (Figures S9–S13). Similar to the base model, the low percentage of patients remaining on trial after ~ 300 days (<20%) resulted in overprediction of tumor size profiles, with observed percentiles being below simulated trends.

Covariate effects and uncertainty in parameter estimates were presented using forest plots, with ratios and 95% CIs constructed using 1000 bootstrap parameter sets over the reference value from the model fit for each fixed covariate effect (Figure S14). Additionally, a forest plot displaying the relative change in tumor shrinkage was constructed using the same method. The reference patient had: three metastatic sites, no liver metastasis, and treatment 53 days after diagnosis. The time duration for the tumor shrinkage calculation was 133 days, which was the approximate time during maintenance when 50% or more of the initial patient population remained in the study. Consistent with previous observations, forest plots showed high uncertainty on Kg/Kd effects. Figure 6 shows simulated tumor profiles over time for the reference patient compared with patients having selected characteristics.

## DISCUSSION

This longitudinal pharmacometric analysis did not identify any significant treatment effects of avelumab versus chemotherapy in the maintenance treatment of advanced GC/GEJC, consistent with the primary analysis of the JAVELIN Gastric 100 trial. Disease models of OS and TGD were developed by integrating covariates efficiently informed by ML methods, and covariates potentially prognostic of OS and TGD were identified. However, no predictive factors associated with OS or TGD during avelumab treatment were found.

The analyses presented provide an example of incorporating ML approaches into a traditional pharmacometric workflow. Specifically, separate RF models were used to identify prognostic factors for OS and tumor size end points, which were then added to parametric TTE and population TGD models, respectively. Supplementing ML with parametric methods resulted in more-interpretable final models than the RF alone, particularly for noisy time-varying covariates and given the multistage trial design (induction and maintenance phases). Furthermore, in comparison to parametric modeling alone and/or using stepwise regression or hypothesis testing, the ML approach was faster and was performed using a single pass over the data. One potential limitation of this workflow arises in the translation of the nonlinear and interacting effects inherent in ML models into parametric forms. We started with linear effects and used diagnostic plots to guide refinement of the model. Alternative approaches could also be considered to guide the initial choice of covariate-effect relationships, such as using partial dependence or accumulated local effect plots.<sup>23</sup>

Most parameters selected by ML exhibited large effects in the parametric model, and those with a smaller effect may also be relevant for future clinical consideration. For several parameters selected by ML, the effect on the estimated mean was small and affected the tail or variance of event times, and not the mean, median, or central tendency. However, considering the number of covariates screened, the plausibility of misspecification from random variability in the data, especially when dealing with smaller covariate effects, cannot be ruled out. Accordingly, the results presented here should be considered hypothesis generating. Nevertheless, the longitudinal models developed using covariates for OS and TGD provide a quantitative framework that can be leveraged as a disease model for GC in the maintenance setting. In conjunction, the parametric model estimated linear relationships, and no model misspecification was evident based on diagnostic plots. Strong, very nonlinear relationships could have been important predictors in ML models for

**TABLE 1** Final structural tumor growth dynamics model parameters and final model variance parameters

	Final model		Nonparametric bootstrap			
	Estimate	Shrinkage	Median	Mean	RSE (%)	95% CI
<b>Structural model parameters</b>						
Kg, 1/year	$\Theta_1$	Tumor growth rate	0.271	NA	11.9	0.249, 0.378
BASE, mm	$\Theta_2$	Baseline tumor size	27.0	NA	5.43	23.9, 29.6
Slope, 1/mm	$\Theta_3$	Slope of proportionality between Kd and Kg	-18.8	NA	115	-50.2, -0.814
Intercept, 1/year*mm	$\Theta_4$	Intercept	5.18	NA	111	0.415, 14.3
Kd, 1/year*mm	$\Theta_3 * \text{Kg} + \Theta_4$	Tumor deceleration rate	0.0852	NA	21.4	0.0756, 0.170
<b>Covariate effect parameters</b>						
$\text{Kg}_{\text{Tdiag}}$	$\Theta_5$	Time from diagnosis effect on Kg-Kd	-0.00291	NA	150	-0.0408, 0.0104
$\text{Kg}_{\text{Livermet}}$	$\Theta_6$	Liver metastasis effect on Kg-Kd	0.0164	NA	111	0.00351, 0.387
$\text{BASE}_{\text{Nmet}}$	$\Theta_7$	No. of metastatic sites effect on BASE	0.143	NA	15.7	0.104, 0.193
$\text{BASE}_{\text{Knon-PD/PR}}$	$\Theta_8$	Re-baseline non-PD/PR effect on BASE	-0.0769	NA	167	-0.249, 0.181
$\text{BASE}_{\text{RSD}}$	$\Theta_9$	Re-baseline stable disease effect on BASE	0.644	NA	20.5	0.423, 0.941
<b>Interindividual variance parameters</b>						
IIV-Kd-Kg	$\Omega_{(1,1)}$		0.00163 [CV% = 4.04]	32.7	131	0.000220, 0.382
IIV-BASE	$\Omega_{(2,2)}$		0.659 [CV% = 96.6]	1.85	5.91	0.585, 0.735
IIV-Kd	$\Omega_1 * \Theta_3^2$		0.576 [CV% = 88.3]	100	30.7	0.204, 0.880
<b>Interindividual covariance parameters</b>						
Kd-Kg	$\Omega_{1,3}$		-0.031 [Corr = 1.00]	NA	92.6	0.0109, 0.429
<b>Residual variance</b>						
Proportional	$\Sigma_{(1,1)}$		0.0505 [CV% = 22.5]	14.8	13.5	0.0365, 0.0624

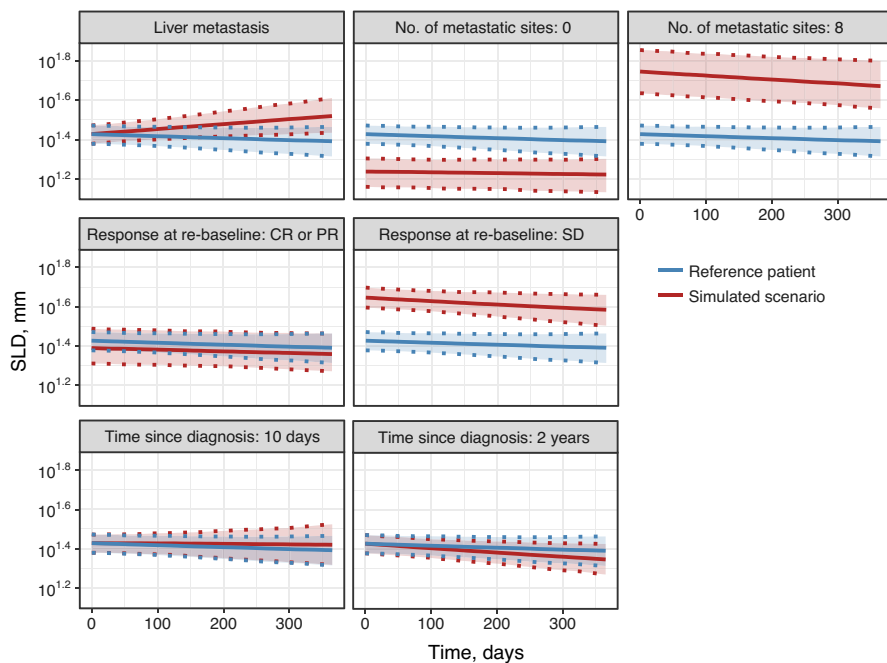
Note: Kd was estimated using Kg, slope, and intercept as follows: Kd = slope\*Kg+intercept.

The CI was determined from the 2.5th and 97.5th percentiles of the nonparametric bootstrap ( $n = 1000$ ) estimates.

Abbreviations: CI, confidence interval; Corr, correlation coefficient; CV, coefficient of variation; IIV, interindividual variance; NA, not applicable; PD, progressive disease; PR, partial response; RSE, relative standard error.

CV% of omega =  $\sqrt{\text{estimate} - 1} * 100$ .

CV% of sigma =  $\sqrt{\text{estimate} * 100}$ .



**FIGURE 6** Simulated tumor profiles over time for identified covariates effects. Blue: Reference patient with three metastatic sites, no liver metastasis, time since diagnosis of 53 days, and re-baseline response of complete response. Red: Simulated scenarios as described in each panel. The 95% prediction intervals using the nonparametric bootstrap results are presented. CR, complete response; PR, partial response; SD, stable disease; SLD, sum of the longest diameter of target lesions

a small subset of patients with extreme values and would have manifested as tail effects.

The assessment of which covariates were stable over time was based on models that evaluated linear trends in time. Although this approach can identify linear and monotonic nonlinear trends, it is possible that longitudinal trends that were nonmonotonic were missed. Based on the study design and relatively sparse collection of data (median number of observations ranged from 4 to 8 across 19 time-varying covariates), the ability to detect nonmonotonic longitudinal trends was limited. Future applications of this methodology should consider the possibility of identifying nonmonotonic longitudinal trends.

Time-invariant (older age, higher GGT levels, absence of peritoneal carcinomatosis, complete or partial response at re-baseline, and re-baseline ECOG PS of 0) and time-varying (lower neutrophil-lymphocyte ratio, lower LDH, lower CRP, and higher albumin) covariates predicting longer OS were identified. Age, CRP, LDH, and neutrophil-lymphocyte ratio have been reported previously as strong prognostic biomarkers in patients with solid tumors.<sup>18,19,24–30</sup> In contrast to our results, GGT has been reported previously as a marker for poor prognosis in patients with GC.<sup>31,32</sup> Clinical trial simulations suggested a benefit with avelumab treatment at milestone survival times greater than 1 year; however, these differences were estimated to be less than 10% and were not considered clinically meaningful, and the probability of exceeding the 10% threshold was small at all landmark times considered. Predicted differences beyond the median OS (e.g., 2 years) were driven primarily through the treatment effect on the log-logistic shape parameter. In the log-logistic model,

the shape parameter influences the variance and, hence, the tails of the survival distribution. The inclusion of a treatment effect on the shape parameter was necessary to characterize the data, even after incorporating the effects of time-varying covariates. This further suggests that no factor included in the parametric model was sufficient to identify a subset of patients likely to survive longer with avelumab versus chemotherapy.

Tumor growth inhibition was also modeled with a combination of parametric models and ML. The analysis was limited by the modest change in tumor size during maintenance treatment and the high percentage of patients who discontinued before median tumor shrinkage in the population data set was observed. Furthermore, because of the shared IIV for tumor growth and deceleration rate in the TGD model, interpretation of some identified effects was complex. The negative slope relating Kg and Kd indicates that effects reducing Kg will increase Kd, with an overall effect on tumor shrinkage that is greater than expected compared with an isolated effect on Kg. Tumor size during maintenance was stable. Although time since diagnosis was identified as a covariate on tumor growth rate constant by ML, simulations from the resulting parametric TGD model that included all ML-identified covariates did not reveal a meaningful association, indicating that the effect was likely not clinically relevant. Baseline characteristics identified by ML as predictors of TGD were liver metastasis for tumor growth and deceleration constant rates. The reduced tumor shrinkage associated with the presence of liver metastasis, which was more marked in the avelumab arm, is consistent with previous reports of reduced efficacy with immunotherapy in patients with liver metastases.<sup>33</sup> Number of metastatic sites and stable

disease at re-baseline were associated with baseline tumor size.

JAVELIN Gastric 100 was a multiregional clinical trial, including countries in the Eastern Asian region (Japan, Republic of Korea, Taiwan, and Thailand) where GC has its highest prevalence.<sup>34</sup> An important finding was that Asian versus non-Asian region was not identified as a covariate in OS or TGD models. Assessment of conservation of disease-related intrinsic and extrinsic factors is an important consideration when applying International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines E5 and E17 for ethnic sensitivity assessment to support Asia-inclusive clinical development strategies.<sup>35–39</sup> The results of our analyses indicate the lack of discernable differences in disease progression or outcomes between Asian and non-Asian populations and are valuable to inform the design of Asia-inclusive trials in GC in the postinduction setting.

In conclusion, our analyses established an innovative workflow supporting ML-enabled pharmacometric modeling of OS and TGD. No significant treatment effect on OS was found within the JAVELIN Gastric 100 population, consistent with the primary analysis; thus, no subpopulation for which avelumab was superior to chemotherapy was identified. However, a disease model for GC in the postinduction/maintenance setting was developed, and potential prognostic factors for both OS and TGD were identified. These require further confirmation but may inform future studies in this setting.

### ACKNOWLEDGEMENTS

The authors would like to thank the patients and their families and the investigators, co-investigators, and study teams at each of the participating centers.

### CONFLICTS OF INTEREST

N. Terranova is an employee of Merck Institute of Pharmacometrics, Lausanne, Switzerland, an affiliate of Merck KGaA, Darmstadt, Germany. J. French reports personal fees from the healthcare business of Merck KGaA, Darmstadt, Germany during the conduct of the study. H. Dai is an employee of EMD Serono. M. Wiens reports personal fees from the healthcare business of Merck KGaA, Darmstadt, Germany during the conduct of the study. A. Khandelwal is an employee of the healthcare business of Merck KGaA, Darmstadt, Germany. A. Ruiz-Garcia reports personal fees from the healthcare business of Merck KGaA, Darmstadt, Germany during the conduct of the study. J. Manitz is an employee of EMD Serono. A. von Heydebreck is an employee of and reports stock ownership in the healthcare business of Merck KGaA, Darmstadt, Germany. M. Ruisi was an employee of EMD Serono when the study was conducted. K. Chin

was an employee of EMD Serono when the study was conducted. P. Girard is an employee of Merck Institute of Pharmacometrics, Lausanne, Switzerland, an affiliate of Merck KGaA, Darmstadt, Germany. K. Venkatakrishnan is an employee of EMD Serono. As an Associate Editor for *CPT: Pharmacometrics & Systems Pharmacology*, Jonathan French was not engaged in the review or decision process for this paper.

### AUTHOR CONTRIBUTIONS

N.T., H.D., J.M., M.W., K.V., J.F., A.R.-G., and A.vH. wrote the manuscript. N.T., M.R., H.D., J.M., P.G., M.W., K.V., J.F., A.R.-G., A.K., A.vH., and K.C. designed the research. N.T., M.R., H.D., M.W., K.V., J.F., A.R.-G., A.vH., and K.C. performed the research. N.T., M.R., H.D., J.M., M.W., J.F., A.R.-G., A.K., A.vH., and K.C. analyzed the data.

### DATA AVAILABILITY STATEMENT

For all new products or new indications approved in both the European Union and the United States after January 1, 2014, the healthcare business of Merck KGaA, Darmstadt, Germany, will share patient-level and study-level data after deidentification, as well as redacted study protocols and clinical study reports from clinical trials in patients. These data will be shared with qualified scientific and medical researchers, upon researcher's request, as necessary for conducting legitimate research. Such requests must be submitted in writing to the company's data sharing portal. More information can be found at <https://www.merckgroup.com/en/research/our-approach-to-research-and-development/healthcare/clinical-trials/commitment-responsible-data-sharing.html>. Where the healthcare business of Merck KGaA, Darmstadt, Germany, has a co-research, co-development, or co-marketing/co-promotion agreement or where the product has been out-licensed, it is recognized that the responsibility for disclosure may be dependent on the agreement between parties. Under these circumstances, the healthcare business of Merck KGaA, Darmstadt, Germany, will endeavor to gain agreement to share data in response to requests.

### REFERENCES

1. Vandeveer AJ, Fallon JK, Tighe R, Sabzevari H, Schlom J, Greiner JW. Systemic immunotherapy of non-muscle invasive bladder cancer with avelumab, an anti-PD-L1 immune checkpoint inhibitor. *Cancer Immunol Res.* 2016;4:452-462.
2. Grenga I, Donahue RN, Lepone LM, Richards J, Schlom J. A fully human IgG1 anti-PD-L1 MAb in an in vitro assay enhances antigen-specific T-cell responses. *Clin Transl Immunol.* 2016;5:e83.
3. Bavencio (avelumab). *Summary of Product Characteristics.* the healthcare business of Merck KGaA, Darmstadt, Germany; 2021.



4. Bavencio (avelumab). *Prescribing Information*. EMD Serono; 2020.
5. Moehler M, Dvorkin M, Boku N, et al. Phase III trial of avelumab maintenance after first-line induction chemotherapy versus continuation of chemotherapy in patients with gastric cancers: results from JAVELIN Gastric 100. *J Clin Oncol*. 2021;39:966-977.
6. Hutchinson L, Steiert B, Soubret A, et al. Models and machines: how deep learning will take clinical pharmacology to the next level. *CPT Pharmacometrics Syst Pharmacol*. 2019;8:131-134.
7. McComb M, Ramanathan M. Generalized pharmacometric modeling, a novel paradigm for integrating machine learning algorithms: a case study of metabolomic biomarkers. *Clin Pharmacol Ther*. 2020;107:1343-1351.
8. Koch G, Pfister M, Daunhawer I, Wilbaux M, Wellmann S, Vogt JE. Pharmacometrics and machine learning partner to advance clinical data analysis. *Clin Pharmacol Ther*. 2020;107:926-933.
9. Sibieude E, Khandelwal A, Hesthaven JS, Girard P, Terranova N. Fast screening of covariates in population models empowered by machine learning. *J Pharmacokinetic Pharmacodyn*. 2021;48(4):597-609.
10. Terranova N, Venkatakrishnan K, Benincosa L. Application of machine learning in translational medicine: current status and future opportunities. *AAPS J*. 2021;23:74.
11. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2:841-860.
12. Lipkovich I, Dmitrienko A, D'Agostino RB Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med*. 2017;36:136-196.
13. Lipkovich I, Dmitrienko A. Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. *J Biopharm Stat*. 2014;24:130-153.
14. Kurska MB, Jankowski A, Rudnicki WR, Boruta – a system for feature selection. *Fundam Inform*. 2010;101:271-285.
15. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med*. 2018;6:121.
16. Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. *CPT Pharmacometrics Syst Pharmacol*. 2012;1:e6.
17. Upton RN, Mould DR. Basic concepts in population modeling, simulation, and model-based drug development: part 3-introduction to pharmacodynamic modeling methods. *CPT Pharmacometrics Syst Pharmacol*. 2014;3:e88.
18. Kambayashi Y, Fujimura T, Hidaka T, Aiba S. Biomarkers for predicting efficacies of anti-PD1 antibodies. *Front Med (Lausanne)*. 2019;6:174.
19. Hiramatsu S, Tanaka H, Nishimura J, et al. Gastric cancer cells alter the immunosuppressive function of neutrophils. *Oncol Rep*. 2020;43:251-259.
20. Kim EY, Song KY. The preoperative and the postoperative neutrophil-to-lymphocyte ratios both predict prognosis in gastric cancer patients. *World J Surg Oncol*. 2020;18:293.
21. Ota Y, Takahari D, Susuki T, et al. Changes in the neutrophil-to-lymphocyte ratio during nivolumab monotherapy are associated with gastric cancer survival. *Cancer Chemother Pharmacol*. 2020;85:265-272.
22. Vaghi C, Rodallec A, Fanciullino R, et al. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS Comput Biol*. 2020;16:e1007178.
23. Molnar C. *Interpretable machine learning. A Guide for Making Black Box Models Explainable*. 2019. <https://christophm.github.io/interpretable-ml-book/> Accessed November 2, 2021.
24. Shrotriya S, Walsh D, Nowacki AS, et al. Serum C-reactive protein is an important and powerful prognostic biomarker in most adult solid tumors. *PLoS One*. 2018;13:e0202555.
25. Hou XM, Yuan SQ, Zhao D, Liu XJ, Wu XA. LDH-A promotes malignant behavior via activation of epithelial-to-mesenchymal transition in lung adenocarcinoma. *Biosci Rep*. 2019;39:BSR20181476.
26. Ding J, Karp JE, Emadi A. Elevated lactate dehydrogenase (LDH) can be a marker of immune suppression in cancer: Interplay between hematologic and solid neoplastic clones and their microenvironments. *Cancer Biomark*. 2017;19:353-363.
27. Theuer CP, de Virgilio D, Keese G, et al. Gastric adenocarcinoma in patients 40 years of age or younger. *Am J Surg*. 1996;172:473-477.
28. Smith BR, Stabile BE. Extreme aggressiveness and lethality of gastric adenocarcinoma in the very young. *Arch Surg*. 2009;144:506-510.
29. Yue L, Lu Y, Li Y, et al. Prognostic value of C-reactive protein to albumin ratio in gastric cancer: a meta-analysis. *Nutr Cancer*. 2021;73:1864-1871.
30. Yu Q, Li K-Z, Fu Y-J, et al. Clinical significance and prognostic value of C-reactive protein/albumin ratio in gastric cancer. *Ann Surg Treat Res*. 2021;100:338-346.
31. Yang S, He X, Liu Y, et al. Prognostic significance of serum uric acid and gamma-glutamyltransferase in patients with advanced gastric cancer. *Dis Markers*. 2019;2019:1415421.
32. Wang Q, Shu X, Dong Y, et al. Tumor and serum gamma-glutamyl transpeptidase, new prognostic and molecular interpretation of an old biomarker in gastric cancer. *Oncotarget*. 2017;8:36171-36184.
33. Yu J, Green MD, Li S, et al. Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. *Nat Med*. 2021;27:152-164.
34. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209-249.
35. Asano K, Aoi Y, Kamada S, Uyama Y, Tohkin M. Points to consider for implementation of the ICH E17 guideline: learning from past multiregional clinical trials in Japan. *Clin Pharmacol Ther*. 2020;109:1555-1563.
36. Venkatakrishnan K, Cook J. Driving access to medicines with a totality of evidence mindset: an opportunity for clinical pharmacology. *Clin Pharmacol Ther*. 2018;103:373-375.
37. Zhou X, Friedlander S, Kupperman E, et al. Asia-inclusive global development of pevonedistat: Clinical pharmacology and translational research enabling a phase 3 multiregional clinical trial. *Clin Transl Sci*. 2021;14(3):1069-1081.
38. Venkatakrishnan K, Burgess C, Gupta N, et al. Toward optimum benefit-risk and reduced access lag for cancer drugs in Asia: a global development framework guided by clinical pharmacology principles. *Clin Transl Sci*. 2016;9:9-22.

39. US Food and Drug Administration E17 General principles for planning and design of multiregional clinical trials guidance for industry. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e17-general-principles-planning-and-design-multi-regional-clinical-trials>. Accessed November 2, 2021.

### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Terranova N, French J, Dai H, et al. Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN Gastric 100 phase III trial of avelumab. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:333-347. doi:[10.1002/psp4.12754](https://doi.org/10.1002/psp4.12754)