Citation: CPT Pharmacometrics Syst. Pharmacol. (2020) 9, 115–123; doi:10.1002/psp4.12489

ARTICLE

Tumor Time-Course Predicts Overall Survival in Non-Small Cell Lung Cancer Patients Treated with Atezolizumab: Dependency on Follow-Up Time

Ida Netterberg[1], René Bruno[2], Ya-Chi Chen[3], Helen Winter[3], Chi-Chung Li[3], Jin Y. Jin[3] and Lena E. Friberg[1,*]

The large heterogeneity in response to immune checkpoint inhibitors is driving the exploration of predictive biomarkers to identify patients who will respond to such treatment. We extended our previously suggested modeling framework of atezolizumab pharmacokinetics, IL18, and tumor size (TS) dynamics, to also include overall survival (OS). Baseline and model-derived variables were explored as predictors of OS in 88 patients with non-small cell lung cancer treated with atezolizumab. To investigate the impact of follow-up length on the inclusion of predictors of OS, four different censoring strategies were applied. The time-course of TS change was the most significant predictor in all scenarios, whereas IL18 was not significant. Identified predictors of OS were similar regardless of censoring strategy, although OS was underpredicted when patients were censored 5 months after last dose. The study demonstrated that the tumor-time course-OS relationship could be identified based on early phase I data.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Cancer immunotherapy with checkpoint inhibitors has revolutionized the cancer treatment landscape. Improved overall survival (OS) has been observed across tumor types, however, the response is highly heterogenic, and it is desirable to evaluate predictors of survival to select patients who are expected to respond to the treatment.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ Relationships among OS and circulating biomarkers, tumor size, pharmacokinetic metrics, and baseline covariates were studied in a parametric time-to-event analysis in 88 patients with non-small cell lung cancer treated with atezolizumab. In addition, four different strategies for censoring OS data were explored.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ The tumor-time course was a predictor of OS, regardless of censoring strategy. None of the evaluated circulating biomarker metrics predicted OS. Included predictors were similar for all four censoring strategies, and earlier data cutoffs could predict survival in longer follow-ups.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ This work shows promise for applying modeling and simulation in oncology to evaluate predictors of OS based on data from phase I.

Overall survival (OS) is considered gold standard for demonstrating clinical benefit in oncology.[1] Population modeling has previously demonstrated usefulness in drug development by identifying relationships between predictors, such as tumor size (TS) dynamics, treatment related and patient characteristics, and OS.[2–4] The experience of using such extensive modeling framework in cancer immunotherapy is, however, limited[5–7] and inclusion of longitudinal biomarker data (other than TS) has not been reported.

The progress in cancer immunotherapy has expanded the therapeutic options for oncology patients and clinical benefit has been observed across a range of different tumor types.[8–10] Atezolizumab is an engineered humanized immunoglobulin G1 monoclonal antibody, currently approved in over 50 countries (including the United States and the European Union) for treatment in patients with advanced urothelial carcinoma and metastatic non-small cell lung cancer (NSCLC), and recently approved by the US Food and Drug Administration (FDA) in patients with triple-negative breast cancer and extensive-stage small-cell lung cancer. Atezolizumab targets the programmed death-ligand 1 (PD-L1), which inactivates the T cell response upon binding to its receptor, programmed death-1, expressed on activated T cells. Tumor cells may express PD-L1 as a mechanism of evading immune destruction,[11,12] and blocking the interaction between PD-L1 and programmed death-1 may sustain the T cell response and increase the antitumor effect. However, the antitumor response is highly heterogenic among patients with cancer

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden; [2]Department of Clinical Pharmacology, Genentech-Roche, Marseille, France; [3]Department of Clinical Pharmacology, Genentech, South San Francisco, California, USA. *Correspondence: Lena E. Friberg (lena.friberg@farmbio.uu.se)
Received: July 11, 2019; accepted: November 4, 2019. doi:10.1002/psp4.12489

treated with checkpoint inhibitors.[13] It is, therefore, desirable to evaluate biomarkers that can be related to clinical benefit. PD-L1 expression is an extensively studied biomarker for treatment with atezolizumab.[14] However, PD-L1 alone does not explain all observed variability in the response.[15–17]

IL-18, a proinflammatory cytokine that stimulates release of interferon-γ from activated T cells,[18] was recently related to TS changes in a pharmacokinetic (PK)/pharmacodynamic modeling framework (PK-IL18-TS model) using a population-based approach.[19] In this analysis, the model-predicted relative change in IL-18 from baseline at day 21 ($RCFB_{IL\text{-}18,d21}$) together with the cycle-specific atezolizumab concentration area under the curve (AUC), were identified as predictors of tumor shrinkage in 88 patients with NSCLC treated with atezolizumab every 3 weeks. Tumor shrinkage may serve as a marker of clinical benefit in oncology. For example, atezolizumab was first granted accelerated approval in urothelial carcinoma in May 2016 by the FDA based on improved objective response rate.[20] However, demonstration of beneficial OS is still needed. Patients are studied either until death or until the study ends, which may include multiple cutoff dates, when OS is the end point. A patient who is still alive at any of the cutoff dates will be right censored (i.e., the time of death is unknown and will occur after the cutoff date). If patients progress before the end of the study, they may discontinue the study treatment and receive another therapy but remain in the study for OS follow-up without further TS evaluation. The results of an OS analysis may consequently be confounded by the cutoff date (censoring time) as well as deaths from other reasons than cancer. It is, therefore, of interest to explore the impact of censoring to handle potential confounding factors.

The previously developed PK-IL18-TS model predicted sustained tumor suppression for high values of $RCFB_{IL\text{-}18,d21}$.[19] This relationship can be associated with a slower on-treatment apparent tumor growth rate, which has been related to longer survival in patients with NSCLC treated with atezolizumab or docetaxel in phase II and III studies.[5] The aim of this analysis was, therefore, to evaluate if a relationship between IL-18 (in combination with other model-predicted variables, such as the tumor time course, as well as baseline covariates) and OS could be established in the same patient population with NSCLC, as previously studied.[19] Furthermore, it was investigated whether the time of censoring had an impact on the final conclusions, by performing the analysis on data using four different censoring strategies.

METHODS

Data

The current analysis was performed on data from 88 patients with relapsed/refractory NSCLC studied in a first-in-human, dose-escalation phase I study, PCD4989g.[21] One patient received 16 doses of 1 mg/kg, whereas the starting dose was 10, 15, and 20 mg/kg in 10, 27, and 50 patients, respectively. Later on, 7 patients among the 88 patients received a total of 139 fixed doses of 1,200 mg. TS was assessed by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1[22] and the sum of longest diameter by computed tomography, which was evaluated every 6 weeks for 24 weeks and thereafter every 12 weeks until disease progression, death, or new systemic treatment. The study was performed in accordance with the Declaration of Helsinki and participants provided written informed consent.

Model development

The previously developed PK-IL18-TS model[19] was first reestimated (TS-related parameters only) with updated covariate values (number of metastatic sites at baseline). Parametric hazard models were subsequently used to describe the time-to-event data in each of the four data sets described below. The exponential, Weibull, Gompertz, log-normal, and log-logistic distributions were explored to describe the distribution of event times, which was followed by identification of predictors in three steps. First, baseline covariates (**Table S1**) were evaluated in a stepwise covariate modeling procedure. Multiple parameterizations were used for covariates based on PD-L1 expression, neutrophil, and lymphocyte count, metastatic sites, race, and smoking. If any of these were included in a forward step, the other parameterizations (**Table S1**) were excluded in the following steps. Second, model-derived variables (**Table 1**) were explored on top of the baseline covariates, using a sequential PK-IL18-TS-OS model. Interferon-inducible T-cell alpha chemoattractant (ITAC) metrics were also explored using an available PK/pharmacodynamic model for ITAC.[19] Model-derived variables were also evaluated stepwise, where the variable that provided the best improvement of the model fit was added to the model first and if any variable improved the model fit significantly, it was added to the model until no more variables provided statistically significant improvement. Third, each included baseline covariate and model-derived predictor was excluded from the model, one by one, until no more could be excluded without resulting in a statistically significant worse model fit. Likelihood ratio tests were used to evaluate statistical significance (P value < 0.05 in both forward and backward steps, corresponding to a objective function value (ΔOFV) of 3.84 for one degree of freedom) of included predictors. Randomization tests were performed to ensure that the actual significance level corresponded to a value equal or close to a ΔOFV of 3.84.[23] The effect of each included baseline covariate and model-derived variable was additive to each other and added exponentially to the baseline hazard ($h_0(t)$). Continuous covariates were included linearly in the exponential domain and centered around the median covariate value. Categorical covariates were added as a relative change from the mode (reference) covariate value. Missing covariate values were imputed with the median covariate value (continuous) or the mode (categorical). The hazard ($h(t)$), including one continuous covariate ($COV_{cont.}$) and one categorical ($COV_{cat.}$) baseline covariate and one model-derived variable ($Variable_{model\text{-}der.}$), can then be summarized with the following equation;

$$h(t) = h_0(t) \times e^{\beta_{cont.} \times (COV_{cont.} - COV_{cont.,median}) + \beta_{cat.} \times COV_{cat.} + \beta_{model-der.} \times Variable_{model-der.}} \quad (1)$$

where $\beta_{cont.}$ is a parameter relating $COV_{cont.}$ to $h(t)$ and $COV_{cont.,median}$ is the median value of $COV_{cont.}$. $\beta_{cat.}$ is a parameter relating $COV_{cat.}$ (which has a value of 0 for the reference category and 1 for the comparing category for

Table 1 Description of evaluated model-derived variables

Variable	Description[a]
$RCFB_{IL-18,d21}$	IL-18 relative change from baseline at day 21, set to 0 until day 21
$RCFB_{ITAC,d21}$	ITAC relative change from baseline at day 21, set to 0 until day 21
AUC_{0-21}-IL-18	IL-18 accumulated change from baseline area under the curve, time-varying until day 21 and then carried forward
AUC_{0-21}-ITAC	ITAC accumulated change from baseline area under the curve, time-varying until day 21 and then carried forward
AUC_{0-21}-Atezolizumab	Atezolizumab accumulated area under the curve, time-varying until day 21 and then carried forward
R_{Growth}	Individual estimate of the individual growth rate
Log R_{Growth}	Log of the individual estimate of the individual growth rate
TSR6	Tumor size ratio at week 6, time-varying until week 6, and then carried forward
TSR12	Tumor size ratio at week 12, time-varying until week 12, and then carried forward
TTG	Time to tumor growth, time-varying until occurrence, and then carried forward
RCFB-TS(t) (extrapolated)	Time course of tumor size relative change from baseline, extrapolated based on EBEs after last observed tumor size measurement
RCFB-TS(t) (carried forward)	Time course of tumor size relative change from baseline, carried forward 3 weeks after last dose
TS(t)-slope (extrapolated)	Current rate of tumor size changes, extrapolated based on EBEs after last observed tumor size measurement
TS(t)-slope (carried forward)	Current rate of tumor size changes, carried forward 3 weeks after last dose

EBE, empirical Bayes estimate; ITAC, interferon-inducible T-cell alpha chemoattractant.
[a]All variables were derived based on all available observed data, also for parameters that were time-varying until a given time.

a dichotomous covariate) to $h(t)$. $\beta_{model\text{-}der.}$ is a parameter relating $Variable_{model\text{-}der.}$ to $h(t)$.

To illustrate relationships between included baseline covariates and OS, a relative hazard (to the typical patient, i.e., assuming median values of included baseline covariates except for the studied covariate) were computed as follows:

$$\text{Relative hazard}_{cont.} = e^{\beta_{cont.} \cdot (COV_{cont.} - COV_{cont.,median})} \quad (1a)$$

$$\text{Relative hazard}_{categorical} = e^{\beta_{cat.} \cdot COV_{cat.}} \quad (1b)$$

The relative hazard is, therefore, 1 if the covariate value is equal to the median covariate value (continuous covariates) or for the reference category (categorical covariates) and values below and above 1 are related to better and worse prognosis, respectively.

Exploration of censoring time

Four different censoring strategies were investigated to explore the impact on inclusion of predictors and the size of their effects, using (i) all available data (AAD), (ii) data censored no later than at a cutoff date set 2 years earlier than in AAD (C2YE), (iii) data censored no later than 2 years after start of treatment for each individual patient (C2YASOT), and (iv) data censored a maximum of 5 months after last dose (C5MALD). The explored censoring strategies were selected arbitrarily to investigate shorter follow-up and influence of postprogression treatment. Models developed for AAD, C2YE, C2YASOT, and C5MALD will here on be referred to as the AAD, C2YE, C2YASOT, and C5MALD models, respectively.

Software

The analysis was performed with NONMEM version 7.4[24] together with the Laplacian estimation method. Population PK-IL18-TS-related parameters were fixed during estimation while the corresponding data were kept in the data set and influenced the estimation of individual PK-IL18-TS parameters during estimation of OS-related parameters. This estimation method is corresponding to the population PK parameters and data method.[25,26] The individual PK-IL18-TS parameters could consequently be affected by the OS data. Model execution and evaluation was supported by functionalities in Perl-speaks-NONMEM[27] version 4.8.9 and the output was handled using R (https://www.R-project.org) and the R-based packages Xpose version 4[20] and ggplot2 version 3.0.0 (www.ggplot2.org). Pirana was used for construction of run-records.[27] Model building was guided by the OFV (i.e., –2·log likelihood), where the ΔOFVs are nominally χ^2 distributed for nested models and the additional number of parameters is the degree of freedom, as well as reasonable relative standard errors (RSEs). RSEs were obtained using the sampling importance resampling approach implemented in Perl-speaks-NONMEM (re-estimated PK-IL18-TS model)[28] or from the NONMEM R covariance matrix (PK-IL18-TS-OS models). Model performance was evaluated with KaplanMeier visual predictive checks (KMVPCs) where the observed data were compared with the 95% confidence interval (CI) computed from 100 simulations. Patients were simulated from their actual date of inclusion until the end of follow-up to account for censoring in the simulations. The adequacy of the included continuous predictors was assessed with KaplanMeier mean covariate visual predictive checks (KMMCVPC),[29] whereas stratified KMVPCs were used to evaluate included categorical covariates.

Prediction of the full data set (AAD)

The value of using the whole follow-up time on the current data set (i.e., AAD) in comparison to the applied censoring times resulting in shorter follow-up was explored by predicting the AAD given the parameter estimates based on final C2YE, C2YASOT, and C5MALD models by setting MAXEVAL = 0 (i.e., no re-estimation) in NONMEM and evaluating by KMVPCs and KMMCVPCs.

RESULTS

Patients and data

Details of the observed PK, biomarker, and TS data have been reported elsewhere.[19] The median time to death was 1.4 years. There were 69 AAD deaths, 56 C2YE deaths, 54 C2YASOT deaths, and 28 C5MALD deaths, and the follow-up times ranged from 16 days to 5.2 (AAD), 3.2 (C2YE), 2 (C2YASOT), and 4.7 (C5MALD) years for the four different censoring strategies. The corresponding KaplanMeier curves, together with a risk table, are available in **Figure S1**. The median time between last dose and end of follow-up was 32 weeks and 32 patients were followed for > 1 year after last dose. Twelve patients still received atezolizumab 2 years after start of treatment.

Re-estimation of PK-IL18-TS model

The data set used here included updated information on number of metastatic sites (which was a covariate for baseline tumor size, sum of longest diameter, in the earlier PK-IL18-TS model) for seven patients. The earlier defined PK-IL18-TS model parameter-covariate relationships remained statistically significant after the re-estimation. A comparison of parameter estimates in the previous and re-estimated models is presented in **Table S2**.

OS models

The exponential distribution (time-constant hazard) was used to describe the distribution of event times in the AAD, C2YE, and C2YASOT models. The Gompertz distribution provided the best model fit for the C5MALD event times. KMVPCs of the final models showed, in general, no misspecification with respect to the distribution of the event times (**Figure 1**). The survival was, however, slightly overpredicted between 1 and 2 years after start of treatment in the C2YE model. The actual significance level was close to the nominal ΔOFV of 3.84, with a minimum ΔOFV = 3.62 and a maximum ΔOFV = 4.79 for the explored baseline covariates. Baseline lymphocyte count (LYM; AAD) and neutrophil/lymphocyte ratio (NLR; C2YE, C2YASOT and C5MALD) were included after the first forward step. NLR provided the next best improvement of the model fit (AAD) in the first forward step, although because NLR was computed partly from LYM, NLR was not included in the next forward steps. LYM/NLR was followed by baseline alkaline phosphatase (ALP) in the second forward step in all four models. PD-L1 expression at baseline (parameterized as PD-L1 + immune cells/tumor mass < 5% or PD-L1 + tumor cells/tumor mass < 50% vs. PD-L1 + immune cells/tumor mass ≥ 5% or PD-L1 + tumor cells/tumor mass ≥ 50%, here on referred to as low and high PD-L1 expression) was included in three models (AAD, C2YE, and C2YASOT) and smoking (parametrized as former/never vs. current) was included in two models (C2YASOT and C5MALD). Two additional baseline covariates (i.e., race (parameterized as white vs. other)) and aspartate aminotransferase were included in the C5MALD model. A summary of all covariates providing a *P* value < 0.05 at each forward step is given in **Table S3**. No covariate was excluded after the backward elimination steps. Lactate dehydrogenase was explored both on normal and log-scale, but did not improve the model fit significantly in any of the four models.

Despite the significance of IL-18 and ITAC as predictors of TS changes, none of the evaluated IL-18 or ITAC variables, or atezolizumab AUC, added predictive value on top of the baseline covariates, whereas all TS-related variables (**Table 1**) resulted in *P* values < 0.05 when tested one at a time. The time course of TS relative change from baseline, carried forward 3 weeks after last dose, RCFB-TS(t) (carried forward), provided the best model fit in one model (AAD), whereas the RCFB-TS(t), extrapolated based on Empirical Bayes estimates after last observed TS measurement provided the best model fit for the other models (i.e., C2YE, C2YASOT, and C5MALD). Introduction of a second model-derived variable resulted in model instability and large RSEs. Therefore, only one model-derived variable was allowed in the final models.

Each included baseline covariate and model-derived predictor were subsequently omitted, one at a time, from the models in a last backward deletion step. PD-L1

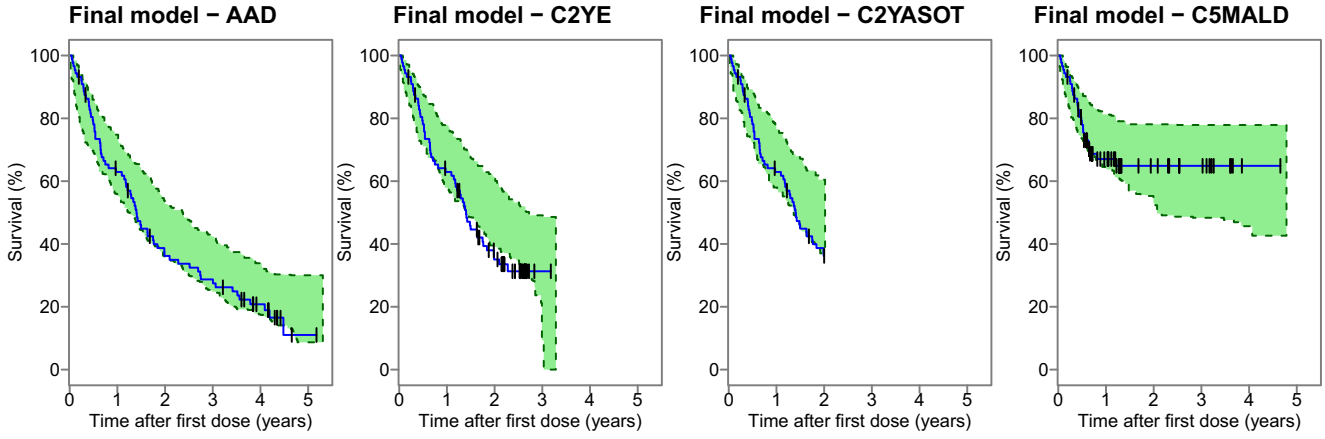

Figure 1 Kaplan–Meier visual predictive checks of the final models for all available data (AAD; first panel), data censored no later than at a cutoff date set 2 years earlier than in AAD (C2YE; second panel), data censored no later than 2 years after start of treatment for each individual patient (C2YASOT, third panel), and data censored a maximum of 5 months after last dose (C5MALD, fourth panel). The plots illustrate the observed Kaplan–Meier curve (blue line) in comparison to the 95% confidence interval, generated from 100 simulations (green shaded area). Black vertical lines indicate censored events.

expression (AAD) and smoking (C2YASOT and C5MALD) were removed in this step. Aspartate aminotransferase and race were further omitted from the C5MALD model to improve model stability because inclusion of these covariates was related to numerical difficulties in the estimation. No predictors were removed from the C2YE model. Included baseline covariates and model-derived predictors after each of the three different steps are summarized **Table 2**. The NONMEM code together with a dummy data set for the final AAD model is provided in **Supplementary Information**. Parameter estimates and corresponding RSEs are presented in **Table 3**. RSEs were in the range of 10–40% for half of all estimated parameters and above 50% for 7 parameters, with the largest RSEs observed in the C5MALD model. The RSEs increased in general the more predictors that were included in the models.

The KMMCVPCs of the base and final AAD models revealed a clear improvement in the final model by capturing the trends (i.e., observed means within the CIs in the final model, as well as providing tighter CIs; **Figure 2**). KMMCVPCs and stratified KMVPCs of the base and final C2YE, C2YASOT, and C5MALD models also illustrated good performance (**Figures S2** and **S3**, respectively) with a tendency of overprediction of the effect of RCFB-TS(t) (extrapolated).

Relationships between included continuous baseline covariates and the relative hazard are illustrated in **Figure 3**. The hazard was lower for lower values of NLR and ALP, where the relative hazard related to the 2.5th percentile of the observed covariate data (i.e., 1.6 and 45 U/L, respectively) ranged between 0.69–0.72 and 0.80–0.84, respectively. The opposite relationship was predicted for LYM and the relative hazard was 1.73 for the 2.5th percentile (i.e., 0.5 10^9/L) of the observed LYM. The median relationships tended to be rather steep at predictions beyond the covariate value corresponding to the 97.5th value (except for LYM), although these predictions were related to wide CIs. Extrapolations of the relative hazard should, therefore, be done cautiously for covariate values larger than the 97.5th percentile. The relative hazard for patients with high PD-L1 expression in comparison to patients with low PD-L1 expression was similar in the C2YE (0.604) and C2YASOT (0.615).

Table 2 Summary of included baseline covariates and model-derived predictors

Data set	Included baseline covariates (first SCM)	Included model-derived predictors	Included in the final model
AAD	LYM ALP PD-L1 expression[a]	RCFB-TS(t) (c-f.)	LYM ALP RCFB-TS(t) (c-f.)
C2YE	NLR ALP PD-L1 expression[a]	RCFB-TS(t) (ext.)	NLR ALP PD-L1 expression[a] RCFB-TS(t) (ext.)
C2YASOT	NLR ALP Smoking[b] PD-L1 expression[a]	RCFB-TS(t) (ext.)	NLR ALP PD-L1 expression[a] RCFB-TS(t) (ext.)
C5MALD	NLR ALP Race[c] Smoking[b] AST	RCFB-TS(t) (ext.)	NLR ALP RCFB-TS(t) (ext.)

AAD, all available data; ALP, alkaline phosphatase; AST, aspartate aminotransferase; C2YE, data censored no later than at a cutoff date set 2 years earlier than in AAD; C2YASOT, data censored no later than 2 years after start of treatment for each individual patient; C5MALD, data censored a maximum of 5 months after last dose; c-f., carried forward 3 weeks after last dose; ext., extrapolated based on empirical Bayes estimates after last observed tumor size measurement; LYM, lymphocyte count; NLR, neutrophil/lymphocyte ratio; PD-L1, programmed death-ligand 1; RCFB-TS(t), time course of tumor size relative change from baseline; SCM, stepwise covariate modeling.
[a]$PD\text{-}L1^+$ immune cells/tumor mass < 5% or $PD\text{-}L1^+$ tumor cells/tumor mass < 50% vs. $PD\text{-}L1^+$ immune cells/tumor mass ≥ 5% or $PD\text{-}L1^+$ tumor cells/tumor mass ≥ 50%. [b]Former/never vs. current. [c]White vs. not white.

Table 3 Final parameter estimates and corresponding RSEs in the final models

	Value (RSE, %)			
	AAD	C2YE	C2YASOT	C5MALD
$Scale_{exp.}$ ($week^{-1}$)	9.69×10^{-3} (12)	8.12×10^{-3} (15)	7.95×10^{-3} (48)	–
$Scale_{Gomp.}$ ($week^{-1}$)	–	–	–	6.39×10^{-3} (56)
$Shape_{Gomp.}$ ($week^{-1}$)	–	–	–	-17.8×10^{-3} (62)
β_{ALP} (L/U)	5.90×10^{-3} (22)	4.84×10^{-3} (44)	4.67×10^{-3} (87)	5.12×10^{-3} (8.4)
β_{LYM} (L/10^9 cells)	-0.780 (37)	–	–	–
β_{NLR}	–	0.154 (31)	0.159 (29)	0.175 (83)
$\beta_{PD\text{-}L1}$	–	−0.505 (63)	−0.486 (67)	–
$\beta_{RCFB\text{-}TS(t)\ (c\text{-}f.)}$	1.44 (18)	–	–	–
$\beta_{RCFB\text{-}TS(t)\ (ext.)}$	–	1.26 (55)	1.32 (17)	1.63 (39)

AAD, all available data; C2YE, data censored no later than at a cutoff date set 2 years earlier than in AAD; C2YASOT, data censored no later than 2 years after start of treatment for each individual patient; C5MALD, data censored a maximum of 5 months after last dose; β_{ALP}, parameter relating alkaline phosphatase to the hazard; β_{LYM}, parameter relating lymphocyte count to the hazard; β_{NLR}, parameter relating the neutrophil/lymphocyte ratio to the hazard; $\beta_{PD\text{-}L1}$, parameter relating high expression of programmed death ligand-1 (parameterized as $PD\text{-}L1^+$ immune cells/tumor mass ≥ 5% or $PD\text{-}L1^+$ tumor cells/tumor mass ≥ 50%) to the hazard; $\beta_{RCFB\text{-}TS(t)\ (c\text{-}f.)}$, parameter relating the time course of tumor size relative change from baseline, carried forward 3 weeks after last dose to the hazard; $\beta_{RCFB\text{-}TS(t)\ (ext.)}$, parameter relating the time course of tumor size relative change from baseline, extrapolated based on empirical Bayes estimates after last observed tumor size measurement to the hazard.
RSEs were computed based on the R covariance matrix in NONMEM.

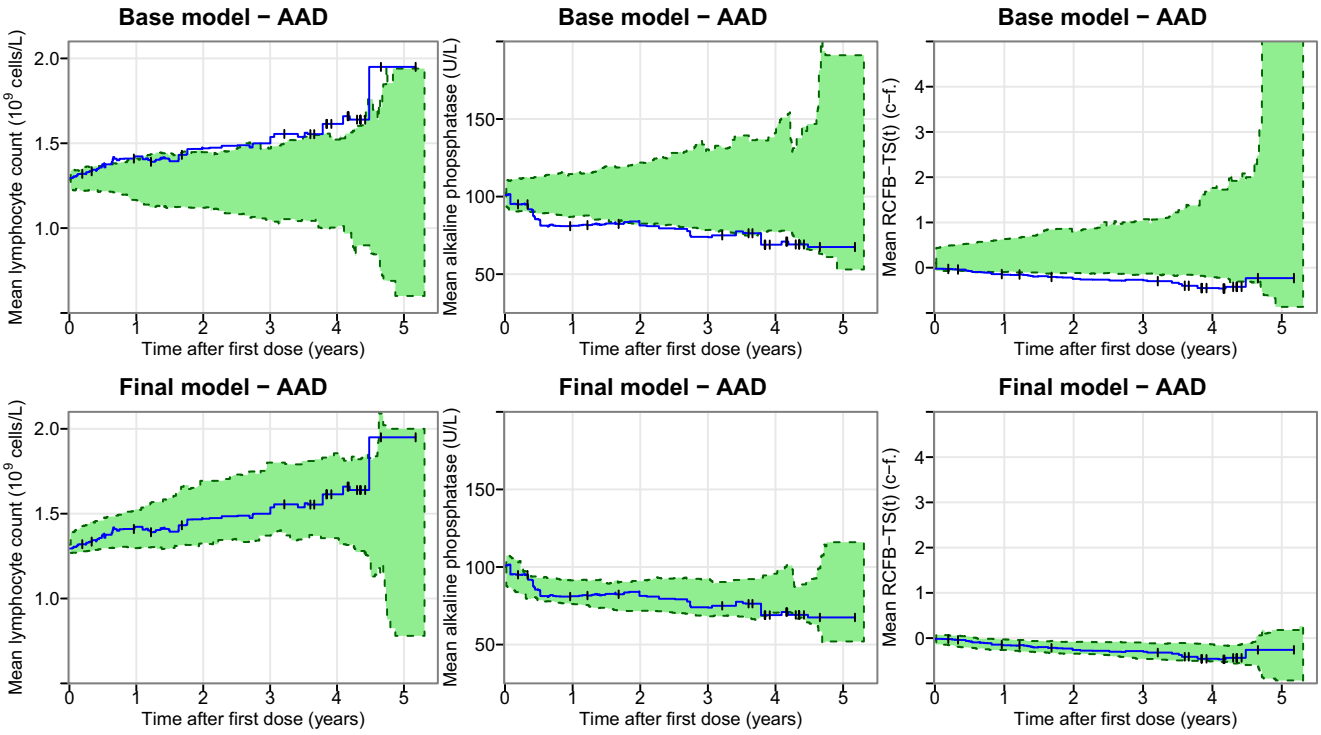

Figure 2 Kaplan–Meier mean covariate visual predictive checks of the base and final models for all available data (AAD). The plots illustrate the observed mean of baseline lymphocyte count (left panel), baseline alkaline phosphatase (middle panel) and time course of tumor size relative change from baseline, carried forward 3 weeks after last dose (RCFB-TS(t) (c-f.); right panel) of patients remaining in the study (blue line) in comparison to the 95% confidence interval, generated from 100 simulations (green shaded area). Black vertical lines indicate censored events.

Prediction of AAD

The KMVPCs illustrated good predictions of the AAD scenario given the final C2YE and C2YASOT models, whereas the final C5MALD model resulted in misspecified predictions (**Figure 4**). However, the misspecification based on the C5MALD model was expected due to the difference in distribution of event times (exponential for AAD and Gompertz for C5MALD) and its lack of long-term survivors after end of treatment. The KMMCVPCs showed in general no apparent misspecification, although the effect of RCFB-TS(t) (extrapolated) was slightly overpredicted when AAD was predicted based on the final C2YE and C2YASOT models (**Figure S4**).

DISCUSSION

The current analysis was performed on data from a phase I study, which contained unusually rich information (e.g., PK, tumor size, biomarker, and OS). In addition, the data provided longer and more matured follow-up information compared with if the analysis had been performed on data from a phase III study at the same time this analysis was performed. Although the number of included patients was limited, we show the potential to use such early data to quantify important relationships between these variables.

In our analysis, the previously suggested PK-IL18-TS framework for atezolizumab[19] was extended to include a relationship with OS. Even though the tumor growth rate, both on normal and log scales (as reported by Claret *et al.* for patients with NSCLC treated with atezolizumab),[5] resulted in model improvement in the current analysis, the RCFB-TS(t) course provided an even better fit for all four models. The RCFB-TS(t) variable carried-forward 3 weeks after last dose performed best in the AAD model, whereas extrapolating the variable based on the empirical Bayes estimates provided the best fit in the C2YE, C2YASOT, and C5MALD models. Our results are comparable to those previously reported in patients treated with checkpoint inhibitors, with respect to relationships between OS and TS-related metrics, where time to tumor growth and current rate of tumor size changes (Tardivon *et al*.), tumor growth rate (Claret *et al*.) and absolute TS time-course and percent change in TS over time (Zheng *et al*.) predicted OS in patients treated with checkpoint inhibitors.[5–7] It is also reasonable to expect that these types of analyses will gain further interest and impact in the future.[30]

IL-18 and AUC were identified as predictors of TS in the earlier PK-IL18-TS analysis, where a high IL-18 response on day 21 after start of atezolizumab treatment was related to a sustained tumor growth inhibition. As mentioned above, the tumor growth rate has been identified as a predictor for OS in a similar group of patients,[5] and the biomarker response may have potential to early assess the benefit of the treatment. In contrast to our hypothesis, none of the IL-18 or ITAC (or atezolizumab AUC) model-derived variables were predictors of OS on top of being a predictor of TS changes.

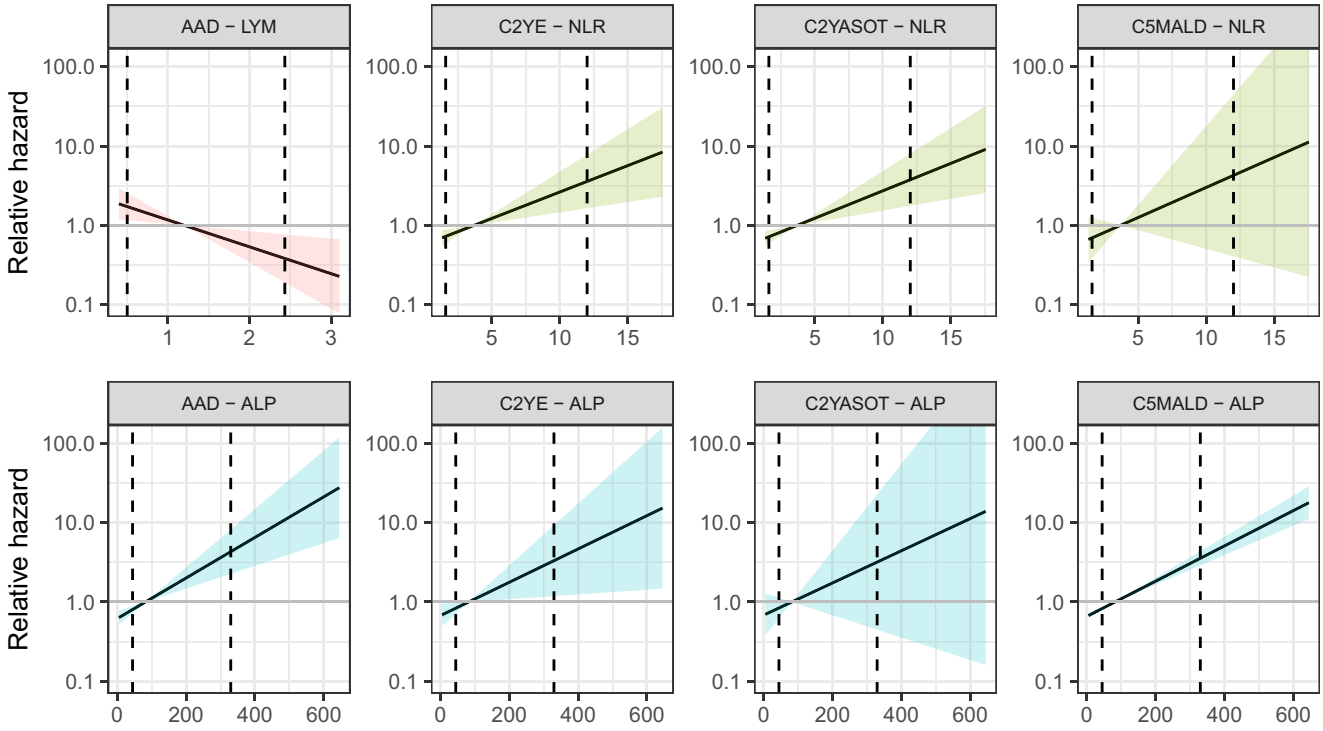

Figure 3 Relative (to the median patient) hazard (logarithmic *y*-axis) illustrated for a change in the continuous baseline covariate value given parameter estimates (solid black lines) in the final model based on all available data (AAD; first panel), data censored no later than at a cutoff date set 2 years earlier than in all available data (C2YE; second panel), data censored no later than 2 years after start of treatment for each individual patient (C2YASOT; third panel), and data censored a maximum of 5 months after last dose (C5MALD; fourth panel). Shaded areas represent the 95% confidence interval, given the corresponding standard error, and are colored by covariate (i.e., pink; lymphocyte count (LYM, 10^9 cells/L), green; neutrophil/lymphocyte ratio (NLR), and blue; alkaline phosphatase (ALP, U/L). The relationships are illustrated for the observed range of the covariate value and dashed vertical lines represent the 2.5th and 97.5th percentiles of the observed covariate value. The horizontal grey line indicates no change in relative hazard (i.e., a patient with median values of all covariates and no change from tumor size baseline).

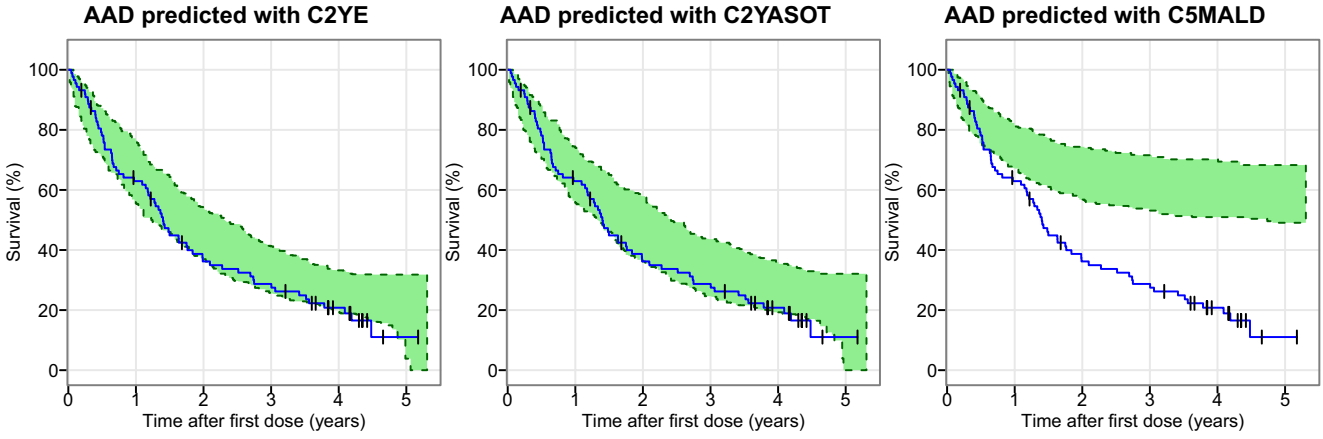

Figure 4 Kaplan–Meier visual predictive checks of all available data (AAD) predicted with data censored no later than at a cutoff date set 2 years earlier than in AAD (C2YE; left panel), data censored no later than 2 years after start of treatment for each individual patient (C2YASOT; middle panel), and data censored a maximum of 5 months after last dose (C5MALD; right panel). The plots illustrate the observed Kaplan–Meier curve (blue line) in comparison to the 95% confidence interval, generated from 100 simulations (green shaded area). Black vertical lines indicate censored events.

Our analysis also addressed the value of long-term follow-up data in phase I (i.e., would conclusions regarding predictors of OS be similar if patients were followed for a shorter time (C2YE and C2YASOT) or if data collected on OS 5 months after they stop taking the study treatment were ignored (C5MALD)). Patients followed for OS are allowed to receive a different treatment, which may confound the results, for example, the median time between last dose and

follow-up was 32 weeks in the current study. Included predictors were overall similar, as well as their corresponding parameter estimates (**Table 3**), among the four different censoring strategies with C5MALD deviating the most.

The current analysis included relatively few patients (n = 88), although it was larger than traditional phase I studies. In addition to that biomarkers were collected, an advantage of the analysis is that the study included more than a single dose level, in contrast to later clinical studies where dose-ranging is less common in oncology. The inclusion of multiple doses reduces the potential correlation between drug exposure and disease status and consequently the selection bias. By allowing the survival data to influence the prediction of the individual TS time course, the impact of immortal time bias is minimized. In addition, similar to other model-based analyses with missing data,[31] it is reasonable to expect that immortal time bias is limited because TS can be predicted at any time point, also after progression.

Inclusion of baseline covariates provided predictive value, where ALP was the only of all explored baseline covariates that was included in all final models. The magnitude of the effect of ALP was similar across all models (4.67×10^{-3}–5.90×10^{-3} L/U). ALP may be elevated in patients with liver disease and may, therefore, reflect patient status.[32] ALP was also a significant predictor of OS in the univariate step in the analysis of phase II and III atezolizumab data in patients with NSCLC.[5]

NLR was included in three final models (C2YE, C2YASOT, and C5MALD) and LYM in the fourth (AAD). Neutrophils promote tumor activity by triggering inflammation in the microenvironment, whereas LYMs are cancer suppressors related to host immunity to cancer.[33,34] It is consequently favorable to have more LYMs relative to neutrophils (i.e., a low NLR). Relationships between NLR and OS have been reported previously across both tumor-types and checkpoint inhibitors.[6,35–37] NLR provided nearly as good improvement as LYM in the AAD model (1.7×10^{-4} vs. 1.0×10^{-4}; **Table S3**) and could consequently have been exchanged to NLR with only minor worsening of the model fit. This was also confirmed in the KMMCVPC for NLR where AAD were well predicted by the models based on C2YE and C2YASOT data (**Figure S4**).

PD-L1 expression was a predictor of OS in the current analysis in the final C2YE and C2YASOT models, where patients with high PD-L1 expression had longer OS in comparison to patients with low PD-L1 expression. PD-L1 expression was explored as a continuous variable and as a categorical variable in various ways but the parameterization as described above consistently provided the best improvement of the model fit in the stepwise covariate modeling (**Table S3**).

By comparing the included predictors in the final models (**Table 2**) it is clear that there were no major differences with respect to censoring strategy. The most significant difference was instead related to the distribution of the event times where the hazard decreased monotonically (Gompertz distribution with a negative shape parameter) in the C5MALD model, whereas it was time-constant in the other three models (exponential distribution). This difference was also obvious when comparing Kaplan–Meier curves (**Figure 1**). The C5MALD Kaplan–Meier curve illustrated a high probability to survive 5 months after last dose, whereas lower survival was reached at the time of censoring in the other three data sets. Regardless of similarities and differences among these four censoring strategies, the current analysis does not address which strategy provides the better option in order to handle potential confounding factors after treatment discontinuation. However, it can be assumed that the C5MALD strategy is the least influenced by other treatments, but survival was overpredicted as discussed above. In addition, a 2-year shorter follow-up, in comparison to using AAD, had only minor impact on the final model with PD-L1 expression included or not, respectively. This was confirmed by predicting the AAD with the C2YE model without major model misspecification (**Figure 4** and **Figure S4**). Given these results, it would have been sufficient to analyze the data earlier and extrapolate to later time points.

An extension to the modeling framework, including atezolizumab PK, IL-18, and TS dynamics, by also including OS was successfully performed using data from the phase I study PCD4989g. The current analysis demonstrated that all explored TS-related model-derived variables improved the model fit, including tumor growth rate, but with RCFB-TS(t) as the best predictor of OS, a finding independent of how the data were censored. Identification of baseline covariates, including LYM/NLR, ALP, and PD-L1 expression, were also independent of censoring time and their estimated effects on OS were similar in all models. It was also shown that the C2YE and C2YASOT models, but not C5MALD, successfully predicted AAD, suggesting that patients could have been followed for a shorter time without loss of important information related to OS. The proposed framework could be used to evaluate similar studies applied in different settings (e.g., other checkpoint inhibitors and tumor types).

Supporting Information. Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (www.psp-journal.com).

Figure S1. Kaplan–Meier curves for the four different censoring strategies together with the corresponding risk table.
Figure S2. Kaplan–Meier mean covariate visual predictive checks of the base and final models.
Figure S3. Kaplan–Meier visual predictive checks stratified by PD-L1 expression in the base and final models.
Figure S4. Kaplan–Meier mean covariate visual predictive checks and Kaplan–Meier visual predictive checks stratified by PD-L1 expression of all available data predicted with the final censored 2 years earlier, censored 2 years after start of treatment and censored 5 months after last dose models.
Table S1. Summary statistics of evaluated baseline covariates.
Table S2. Comparison of parameter estimates in the previous and re-estimated PK-IL18-TS models.
Table S3. Summary of all covariates providing a P value < 0.05 in the baseline covariate stepwise covariate modeling.
Supinfo. Example data set and NONMEM control stream for the AAD PK-IL18-TS-OS model.

Funding. The performed analysis was supported by Genentech Inc. and the Swedish Cancer Society.

Conflict of Interest. I.N. was supported by a grant from Genentech Inc. I.N. and L.F. have acted as paid consultants to Genentech Inc. for other projects. R.B., Y.C., H.W., C.-C.L., J.-Y.J. are Genentech employees. As Deputy Editor-in-Chief for *CPT: Pharmacometrics & Systems Pharmacology*, Lena Friberg was not involved in the review or decision process for this paper.

Author Contributions. I.N., R.B., Y.C., H.W., C.C.L., J.Y.J., and L.E.F. wrote the manuscript. I.N., R.B., and L.E.F. designed the research. I.N., R.B., Y.C., H.W., C.C.L., J.Y.J., and L.E.F performed the research. I.N. and L.E.F. analyzed the data.

1. US Department of Health and Human Services, Food and Drug Administration, Oncology Center of Excellence, Center for Drug Evaluation and Research (CDER) & Center for Biologics Evaluation and Research (CBER) Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. Guidance for Industry. (2018).
2. Claret, L. *et al.* Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J. Clin. Oncol.* **27**, 4103–4108 (2009).
3. Claret, L. *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.* **31**, 2110–2114 (2013).
4. Bender, B.C., Schindler, E. & Friberg, L.E. Population pharmacokinetic-pharmacodynamic modelling in oncology: a tool for predicting clinical response. *Br. J. Clin. Pharmacol.* **79**, 56–71 (2015).
5. Claret, L. *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.* **24**, 3292–3298 (2018).
6. Zheng, Y. *et al.* Population modeling of tumor kinetics and overall survival to identify prognostic and predictive biomarkers of efficacy for durvalumab in patients with urothelial carcinoma. *Clin. Pharmacol. Ther.* **103**, 643–652 (2018).
7. Tardivon, C. *et al.* Association between tumor size kinetics and survival in urothelial carcinoma patients treated with atezolizumab: implication for patient's follow-up. *Clin. Pharmacol. Ther.* **106**, 810–820 (2019). https://doi.org/10.1002/cpt.1450
8. Balar, A.V. & Weber, J.S. PD-1 and PD-L1 antibodies in cancer: current status and future directions. *Cancer Immunol. Immunother.* **66**, 551–564 (2017).
9. Marconcini, R. *et al.* Current status and perspectives in immunotherapy for metastatic melanoma. *Oncotarget* **9**, 12452–12470 (2018).
10. Saudemont, A., Jespers, L. & Clay, T. Current status of gene engineering cell therapeutics. *Front. Immunol.* **9**, 153 (2018).
11. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
12. Chen, D.S. & Mellman, I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39**, 1–10 (2013).
13. Chen, D.S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
14. Patel, S.P. & Kurzrock, R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol. Cancer Ther.* **14**, 847–856 (2015).
15. Rittmeyer, A. *et al.* Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet Lond. Engl.* **389**, 255–265 (2017).
16. Garon, E.B. *et al.* Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
17. Grigg, C. & Rizvi, N.A. PD-L1 biomarker testing for non-small cell lung cancer: truth or fiction? *J. Immunother. Cancer* **4**, 48 (2016).
18. Novick, D., Kim, S., Kaplanski, G. & Dinarello, C.A. Interleukin-18, more than a Th1 cytokine. *Semin. Immunol.* **25**, 439–448 (2013).
19. Netterberg, I. *et al.* A PK/PD analysis of circulating biomarkers and their relationship to tumor response in atezolizumab-treated non-small cell lung cancer patients. *Clin. Pharmacol. Ther.* **105**, 486–495 (2019).
20. Ning, Y.-M. *et al.* FDA approval summary: atezolizumab for the treatment of patients with progressive advanced urothelial carcinoma after platinum-containing chemotherapy. *Oncologist* **22**, 743–749 (2017).
21. Herbst, R.S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).
22. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer Oxf. Engl.* **45**, 228–247 (2009).
23. Wählby, U., Jonsson, E.N. & Karlsson, M.O. Assessment of actual significance levels for covariate effects in NONMEM. *J. Pharmacokinet. Pharmacodyn.* **28**, 231–252 (2001).
24. Beal, S., Sheiner, L., Boeckmann, A. & Bauer, R. (eds) NONMEM 7.4 Users Guides. ICON plc, Gaithersburg, MD (1989–2018).
25. Wade, J. & Karlsson, M.Combining PK and PD data during population PK/PD analysis. Abstract 139, PAGE 8 (1999). <https://www.page-meeting.org/default.asp?abstract=139>
26. Zhang, L., Beal, S.L. & Sheiner, L.B. Simultaneous vs. sequential analysis for population PK/PD data I: best-case performance. *J. Pharmacokinet. Pharmacodyn.* **30**, 387–404 (2003).
27. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacomet. Syst. Pharmacol.* **2**, e50 (2013). https://doi.org/10.1038/psp.2013.24.
28. Dosne, A.-G., Bergstrand, M. & Karlsson, M.O. An automated sampling importance resampling procedure for estimating parameter uncertainty. *J. Pharmacokinet. Pharmacodyn.* **44**, 509–520 (2017).
29. Hooker, A. & Karlsson, M.The Kaplan-Meier Mean Covariate plot (KMMC): a new diagnostic for covariates in time-to-event models. Abstract 2564, PAGE 21 (2012). <https://www.page-meeting.org/default.asp?abstract=2564>
30. US Food & Drug Administration (FDA) & International Society of Pharmacometrics (IsoP) FDA-ISoP Public Workshop: Model Informed Drug Development (MIDD) for Oncology Products. (2018).
31. Gastonguay, M.R. *et al.* Missing data in model-based pharmacometric applications: points to consider. *J. Clin. Pharmacol.* **50**, 63S–74S (2010).
32. Van Hoof, V.O. & De Broe, M.E. Interpretation and clinical significance of alkaline phosphatase isoenzyme patterns. *Crit. Rev. Clin. Lab. Sci.* **31**, 197–293 (1994).
33. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* **454**, 436–444 (2008).
34. Gooden, M.J.M., de Bock, G.H., Leffers, N., Daemen, T. & Nijman, H.W. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br. J. Cancer* **105**, 93–103 (2011).
35. Bagley, S.J. *et al.* Pretreatment neutrophil-to-lymphocyte ratio as a marker of outcomes in nivolumab-treated patients with advanced non-small-cell lung cancer. *Lung Cancer Amst. Neth.* **106**, 1–7 (2017).
36. Ogata, T. *et al.* Neutrophil-to-lymphocyte ratio as a predictive or prognostic factor for gastric cancer treated with nivolumab: a multicenter retrospective study. *Oncotarget* **9**, 34520–34527 (2018).
37. Cassidy, M.R. *et al.* Neutrophil to lymphocyte ratio is associated with outcome during ipilimumab treatment. *EBioMedicine* **18**, 56–61 (2017).