Citation: CPT Pharmacometrics Syst. Pharmacol. (2019) 8, 825–834; doi:10.1002/psp4.12454

ARTICLE

Linking Tumor Growth Dynamics to Survival in Ipilimumab-Treated Patients With Advanced Melanoma Using Mixture Tumor Growth Dynamic Modeling

Yan Feng[1,*], Xiaoning Wang[2], Satyendra Suryawanshi[1], Akintunde Bello[1] and Amit Roy[1]

Early tumor assessments have been widely used to predict overall survival (OS), with potential application to dose selection and early go/no-go decisions. Most published tumor dynamic models assume a uniform pattern of tumor growth dynamics (TGDs). We developed a mixture TGD model to characterize different patterns of longitudinal tumor sizes. Data from 688 patients with advanced melanoma who received ipilimumab 3 or 10 mg/kg every 3 weeks in a phase III study (NCT01515189) were used in a TGD-OS analysis. The mixture model described TGD profiles using three subpopulations (no-growth, intermediate, and fast). The TGD model showed a positive exposure/dose-response (i.e., a higher proportion of patients in no/intermediate growth subpopulations and a lower tumor growth rate with ipilimumab 10 mg/kg relative to the 3 mg/kg dose). Finally, the mixture TGD model-based measures of tumor response provided better predictions of OS compared with the nonmixture model.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Early tumor assessments can predict overall survival (OS); however, a tumor growth dynamic (TGD) model linking tumor response with OS has not been reported for ipilimumab.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ This analysis characterized the association between measures of tumor response derived from a mixture TGD model and OS with ipilimumab in patients with advanced melanoma from a randomized phase III study.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ The mixture TGD model showed better description of individual tumor data than a nonmixture model using three (no-growth, intermediate, and fast) subpopulations. The TGD model showed that patients with ipilimumab 10 mg/kg had a lower tumor growth rate and had more patients in no-growth and intermediate subpopulations than those with ipilimumab 3 mg/kg.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ A mixture TGD model allows accurate estimation of different patterns of tumor dynamics and characterization of the dose/exposure-response relationship. Because precise estimation of individual tumor profiles is critical for long-term OS prediction, this may be used for dose selection and go/no-go decisions.

Tumor growth dynamic (TGD) modeling linking tumor response to overall survival (OS) has enormous potential to inform early go/no-go decision making and dose selection in oncology drug development. However, because it is underutilized, a robust characterization of the tumor response-OS relationship remains elusive. Unlike conventional Response Evaluation Criteria in Solid Tumors (RECIST) in clinical anticancer therapy, the TGD model describes the entire time profile of the tumor response and offers the potential to use higher-resolution tumor burden (TB) data to predict OS. Clinical TGD models[1–5] have been reviewed extensively[6,7] to describe the effects of chemotherapy or targeted therapy in solid tumors. By linking certain tumor size metrics to OS, these analyses have extended beyond modeling tumor size, making it possible to predict the efficacy of long-term treatment.

Prior studies have investigated TGD modeling of anticancer agents, including the immunotherapeutic agents ipilimumab (anticytotoxic T-lymphocyte antigen-4) and nivolumab (anti-programmed death 1).[8] Several studies have demonstrated that early tumor response is a predictor of OS.[8–10] Tumor dynamics were modeled together with OS to identify biomarkers of efficacy in patients with durvalumab-treated urothelial carcinoma.[11] TB mixture modeling data were also reported in pembrolizumab-treated patients with melanoma.[12] Although each individual's TB-time profile can differ by estimated interindividual variability in model parameters, the reported models with a unimodal parameter distribution

[1]Clinical Pharmacology and Pharmacometrics, Bristol-Myers Squibb, Princeton, New Jersey, USA; [2]Metrum Research Group, Tariffville, Connecticut, USA. *Correspondence: Yan Feng (yan.feng@bms.com)
Received: March 22, 2019; accepted: June 27, 2019. doi:10.1002/psp4.12454

do not adequately account for the heterogeneity in tumor response. Therefore, it is important to identify distinct patterns of longitudinal tumor size-time profiles and potentially link these patterns to treatment or patient characteristics. An appropriate approach to account for the heterogeneity in tumor response is with mixture models that describe differentiated TGD patterns as subpopulations. In a general finite mixture model, the total number of subpopulations is predefined, and the subpopulation to which each individual observation belongs is estimated as a latent variable. Advantages of mixture modeling and its clinical applications have been described elsewhere.[13–15]

The relation between treatment and response, either efficacy or safety, has been described to a lesser extent by mixture modeling despite the heterogeneous nature of such relationships. Tumor size data were initially modeled from pembrolizumab-treated patients with melanoma using a mixture model with four subpopulations.[12] This study demonstrated that both the initial mixture model and the later consolidated tumor-size model were able to successfully describe the longitudinal tumor kinetics.[12] However, the relationship between the characterized tumor growth (TG) and survival benefit was not reported.

Ipilimumab is a fully human, immunoglobulin G1 monoclonal antibody that blocks the immune-checkpoint target CTLA-4, which, in turn, leads to T-cell activation resulting in tumor cell death.[16,17] Ipilimumab was the first immune-checkpoint inhibitor to show an improvement in OS of patients with metastatic melanoma.[18,19] It is approved in several countries at 3 mg/kg, given every 3 weeks, for four doses. The benefit-risk of ipilimumab 3 vs. 10 mg/kg was evaluated in patients with advanced melanoma in phase II (CA184-022)[20] and phase III (CA184-169, NCT01515189)[21] studies. In patients with advanced melanoma, ipilimumab 10 mg/kg resulted in significantly longer OS compared with ipilimumab 3 mg/kg (hazard ratio (HR), 0.84; 95% confidence interval (CI), 0.70–0.99; $P = 0.04$).[21] However, the objective response rate (ORR) by RECIST criteria, duration of response, and progression-free survival (HR, 0.89; 95% CI, 0.76–1.40; $P = 0.16$) were similar for both doses.

Here, we have utilized TB data from the CA184-169 study[21] to investigate a TGD mixture modeling approach. We hypothesized that an expanded mixture model linking TGD to OS would be able to outperform the nonmixture model by better describing the heterogeneity in TB. Assigning each individual TGD observation to a subpopulation would facilitate investigation of factors that influence patterns of tumor response to long-term treatment effect, such as OS. A successful classification of several distinct patterns of tumor size-change profiles via mixture modeling following ipilimumab treatment can enable identification of one or more features of tumor response not reflected in ORR categorization by RECIST, but which may be associated with OS. In addition, we hypothesized that the effect of drug exposure would be entirely reflected in tumor response and, in turn, OS.

METHODS

CA184-169 study design

The time course of TB in patients with previously treated or untreated advanced melanoma in the phase III study CA184-169[21] was described by a nonlinear mixed-effects TGD model. In both 3 and 10 mg/kg dose groups, ipilimumab was administered by intravenous infusion for 90 minutes every 3 weeks for four doses. The analysis included data from 688 patients who received ipilimumab ($n = 343$ for 3 mg/kg and $n = 345$ for 10 mg/kg) and for whom TB data were available. The sum of the longest diameters of target lesions based on immune-related response criteria was used as a surrogate for TB. The protocol-specified tumor assessments were at weeks 12, 16, and 24.

Patient eligibility criteria and the treatment design for study CA184-169 have been reported previously.[21] The study was approved by institutional review boards and independent ethics committees at participating institutions and was carried out in accordance with the ethical principles of the Declaration of Helsinki.

TGD model

TGD model development included two steps: (i) a base TGD model determined the most appropriate structural model; and (ii) an exposure-response model evaluated the effects of covariates and exposure on the structural model parameters.

The base model characterized the differences in patterns of response to immune-checkpoint inhibitor therapy. Three models evaluated: (i) a nonmixture model with unimodal distributions for TGD model parameters; (ii) a mixture model with two subpopulations; and (iii) a mixture model with three subpopulations. The nonmixture model was the same as a previously reported model[1] developed with data from cytotoxic agents, which has also been used to describe data from an immunotherapeutic agent.[8] This model was modified to improve the description of TB time profiles that asymptotically approach a steady-state value—an immunotherapy-specific response pattern observed in some patients.

The TB at a given time (t) in the mixture model was described by the following structural model for each population:

Subpopulation 1 (fast TG):

$$TB_i(t) = TB_0 \times e^{-TS_i t} + TG_i \times t$$

Subpopulation 2 (no-growth):

$$TB_i(t) = TB_0 \times e^{-TS_i t} + TB_{SS_i}$$

Subpopulation 3 (intermediate TG and tumor shrinkage (TS)):

$$TB_i(t) = TB_0 \times e^{-TS_i t} + TG_i \times t$$

where $TB_i(t)$ is the TB at time t for the i^{th} patient, and $TB0_i$, TS_i, and TG_i represent baseline TB, TS rate constant, and linear TG rate for the ith patient, respectively.

The original model was modified by the addition of TB_{SS_i}, to describe steady-state TB for the i^{th} patient. TB_{ss} was set to zero in the nonmixture model, which, therefore, collapsed to the original Wang model. TB_{ss} or TG rate was set to zero to describe mixture model subpopulations. Each of the mixture models included one subpopulation to describe the no-growth subpopulation (TG fixed to zero). The Bayesian information criterion (BIC) was used to guide

model selection. Interindividual variability in the structural model parameters was described by the following equation:

$$P_i = P_{TV} \times e^{\eta_{p,i}}$$

where P_i is the value of a TGD structural model parameter (TB0, TS, TG, or TB_{SS}) for the i^{th} patient, P_{TV} is the typical value of P for the population, and $\eta_{p,i}$ is a random realization from a normally distributed random variable with a mean of zero and variance of ω_P^2 that describes the deviation of P_i from P_{TV}.

The difference between observed values and the corresponding model-predicted values was described by a combined residual errors analysis (proportional and additive) and is given by:

$$TB_{i,o}(t) = TB_i(t) \times (1 + \theta_{PROP} \times \varepsilon_{ij}) + \theta_{ADD} \times \varepsilon_{ij}$$

where $TB_{i,o}(t)$ and $TB_i(t)$, respectively, are the observed and model-predicted TB at time t for the ith patient; $\varepsilon_{ij} \sim N(0,\sigma^2)$ is a normally distributed random variable with a mean of zero and variance (σ^2) of one; and θ_{PROP} and θ_{ADD}, respectively, are SDs of proportional and additive components of the residual error.

After developing the base model, the exposure-response model was developed by estimating the effect of time-averaged concentration after the first dose (Cavg1) on the structural model parameters of the TGD model (except baseline tumor burden (TB0)). In addition, baseline lactate dehydrogenase (LDH) was included as a covariate effect. The LDH effect on TB0 was found to be significant in a previous analysis.[9] Baseline clearance (CL) was hypothesized as a surrogate biomarker to reflect patient disease status.[1,22] Therefore, ipilimumab CL was also evaluated in a sensitivity analysis. Patient-specific values of ipilimumab CL and Cavg1 were obtained from a population pharmacokinetics analysis.[22]

The following measures of tumor response were determined using the TGD model: TG, TS, relative change in TB at week 8 (CTB8), and progression rate at week 8 (PRW8). CTB8 was calculated as the TB at week 8 (percentage of TB0), and PRW8 was calculated at the first derivative of TB relative to time at week 8, which represents the slope effect. The effects of these measures of tumor response on OS were evaluated in the subsequent analysis.

OS model

The time-to-death event was described by a semiparametric Cox proportional hazards (CPH) model, with the hazard of death expressed as:

$$\lambda(t) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{X_i})$$

where $\lambda_0(t)$ is the baseline hazard function and $\boldsymbol{X_i}$ is a vector of predictor variables. The parameter vector $\boldsymbol{\beta}$ was estimated by maximum partial likelihood.

The objectives of the OS analysis were to (i) determine the measure of tumor response that provided the best description of OS, and (ii) assess the importance of refinements in the TGD model based on OS. The effects of tumor response measures on OS were evaluated using a full modeling approach considering all other covariate effects on OS simultaneously. Evaluated variables in the full model were baseline TB, age, body weight, sex, Eastern Cooperative Oncology Group (ECOG) status, M stage, LDH, missing postbaseline TB (yes/no), number of target lesions at baseline, and measures of tumor response.

The CPH model was developed in two stages. First, the best measure of tumor response was selected from among four alternative measures (TG, TS, PRW8, and CTB8) by assessing these measures in a CPH model that included all of the covariates listed above. The tumor response measure that provided the lowest BIC was selected for the subsequent model evaluations. An evaluation of the suitability of the selected tumor response to serve as a sufficient statistic for the treatment effect was conducted by estimating the treatment (3 vs. 10 mg/kg) effect with or without the effect of tumor response in the model. Second, a sensitivity analysis was conducted to evaluate the added value of the mixture TGD model on the description of OS by (i) comparing the goodness-of-fit of CPH models with the effect of the tumor response derived from mixture and nonmixture models, and (ii) determining the effect of the mixture subpopulation as a predictor variable in the CPH model of OS.

The CPH model fitting was evaluated by a visual predictive check (VPC) comparing the model-predicted cumulative time-to-event distributions with the corresponding Kaplan-Meier (K-M) curve of the observed data. The CPH model-predicted cumulative probability of death for each individual was used to simulate the occurrence of events and subsequently calculate the cumulative time-to-event distribution using the K-M analysis. There were 1,000 such simulations performed to construct the 90% prediction intervals (PIs) of the distribution. The K-M curve of the observed data was overlaid on the PIs.

The TGD model was developed with the NONMEM computer program version 7.3 (ICON Development Solutions, Hanover, MD), compiled using Intel Fortran. The OS model was developed and model diagnostic plots were generated using the R software version 3.0.2.

RESULTS

TGD model

Patient demographics were similar for both groups in the analysis population (**Table 1**). The base model with three subpopulations (TGD-model 3) had the best fit to the observed data as indicated by the lowest BIC value (**Table S1**). Adding the effect of Cavg1 on TG (TGD-model 4, TGD-model 8, and TGD-model 11) lowered BIC values in models compared with those with Cavg1 effect on TS (TGD-model 5, TGD-model 9, and TGD-model 12). When Cavg1 effect on both TG and TS was included, the uncertainty of parameter estimates substantially increased, indicating overparameterization. Based on the results of exposure-response model development, a mixture model with three subpopulations, Cavg1 effect on TG, and LDH effect on TG/TB0 provided the best fit of data (TGD-model 11). Hence, this model was used to estimate measures of tumor response in the subsequent OS model development. The parameter estimates from covariate TGD mixture model are presented in **Table 2**. Estimated LDH effects on TB0 and TG showed

Table 1 Patients demographics in the analysis population

Characteristic	3 mg/kg (*n* = 343)	10 mg/kg (*n* = 345)	All (*N* = 688)
Mean age (SD), years	60.9 (13.3)	59.0 (14.6)	59.9 (14.0)
Mean body weight (SD), kg	79.3 (17.4)	80.6 (17.7)	79.9 (17.6)
Mean baseline TB (SD), cm	10.0 (8.5)	9.3 (9.0)	9.6 (8.7)
Median LDH ratio[a] (range)	1.0 (0.4–29.4)	1.0 (0.5–40.5)	1.0 (0.4–40.5)
ECOG status, *n* (%)			
0	237 (69.1)	248 (71.9)	485 (70.6)
≥ 1	106 (30.9)	97 (28.1)	203 (29.5)
Sex, *n* (%)			
Male	222 (64.7)	209 (60.6)	431 (62.6)
Female	121 (35.3)	136 (39.4)	257 (37.4)
M stage, *n* (%)			
M0 or M1a	59 (17.2)	58 (16.8)	117 (17.0)
M1b	75 (21.9)	69 (20.0)	144 (20.9)
M1c	209 (60.9)	218 (63.2)	427 (62.1)

ECOG, Eastern Cooperative Oncology Group; LDH, lactate dehydrogenase; TB, tumor burden.
[a]LDH ratio indicates patient's actual value divided by the upper limit of normal. Log-transformed LDH ratio was used in tumor growth dynamics and overall survival model development due to skewed distribution.

that higher baseline LDH was associated with a higher TG rate and higher baseline tumor size. TG decreased with Cavg1; however, the 95% CI of parameter estimate included zero. In addition, compared with the ipilimumab 3 mg/kg arm, the ipilimumab 10 mg/kg arm had a higher percentage of patients in the no-growth group (28.7% vs. 24.2%) and a lower fraction in the fast TG group (40.9% vs. 46.9%; **Table S2**). The ipilimumab treatment effect (10 vs. 3 mg/kg) on TG was assessed in TGD-model 7 nonmixture model development (**Table S1**), but the BIC value was high relative to the TGD model with Cavg1 (TGD-model 6).

For the mixture model with three subpopulations, patients were categorized into no-growth, intermediate TG and TS, and fast TG groups. The no-growth group included patients with TS leading to a constant steady-state TB, the intermediate group included patients with initial shrinkage followed by tumor progression, and the fast TG group included patients with fast tumor progression early on during the study. The three subpopulations were determined by the mixture model identifying patients with qualitatively different TGD, as evidenced by differences in the distributions of TG and TS values. TB0 was higher in patients with fast TG, and TS was higher in patients with no-growth compared with other subpopulations. Patients in the fast TG subpopulation had higher TG relative to patients in the intermediate TG and TS group (**Figure 1**).

The observed and model-predicted time course of TB profile showed three patterns of tumor dynamics. Based on model evaluations of the exposure-response TGD model, there was good agreement between the observed (**Figure 2a**) and predicted (**Figure 2b**) time course of change of TB from baseline as well as the observed vs. model-predicted change of TB from baseline at the first tumor assessment at week 12 (**Figure S1a,b**). Additional sensitivity analyses (TGD-model 15) showed that BIC was higher in the mixture model relative to the final TGD model (TGD-model 11), indicating that TB_{ss} reflected the tumor pattern in the subpopulation. Results of the other sensitivity analyses showed that although ipilimumab baseline CL was a prognostic factor for OS, the inclusion of CL in mixture model TGD-model 14 did not improve the BIC (**Table S1**) in this model that also included Cavg1.

OS model

The effects of the following tumor response measures derived from the final TGD model were evaluated in a CPH model that included prespecified covariate effects: CTB8, TG, TS, and PRW8. The model development started with model 0, which only included a prespecified covariate without including tumor response measures. OS-model 1 with PRW8 from the covariate mixture model (PRW8.MIXC) had the lowest BIC value, relative to other measures of tumor response—OS-model 0 without including PRW8 and model 5 including the effect of objective response (responder (best overall response of complete response/partial response) vs. nonresponder; **Table S3**). Therefore, PRW8.MIXC was selected and evaluated further by sensitivity analysis. **Figure 3** shows all the estimated effects in the CPH model (OS-model 1) and the HRs of OS across the predictor ranges along with 95% CIs. The predictor variables with a significant effect on OS were ECOG status, M stage (M1C vs. M1A/M0), baseline TB only (yes or no), baseline tumor size, and baseline LDH (95% CI did not include 1). The risk of death increased with an increase in PRW8.MIXC, LDH, and baseline tumor size. The risk of death was also higher in patients with M stage of M1C, with ECOG = 1, and who had baseline TB only, indicating the early drop-off might be associated with lack of efficacy, which is in agreement with observed high-risk data in patients with nonevaluable objective response (**Figure S2**).

The first sensitivity analysis was conducted to evaluate model performance using PRW8.MIXC and PRW8 derived from the nonmixture model (PRW8.NMIXC). The OS-model S1 with PRW8.NMIXC had a 33-unit higher BIC relative to OS-model 1 with PRW8.MIXC (**Table S3**). The second sensitivity analysis was performed to evaluate treatment

Table 2 Parameter estimates of TGD mixture model with three subpopulations

Parameter[a,b]	Estimate[c]	95% CI[d]
Fixed effects		
No growth		
TB0 (cm)	2.53	2.05–3.00
TS (1/week)	0.0458	0.0340–0.0576
TB_{ss}	1.71	1.15–2.26
Intermediate		
TB0 P3 (cm)	5.83	4.69–6.98
TG P3 (cm/week)	0.0236	0.00746–0.0398
TS P3 (1/week)	0.00299	7.56E-04 to 0.00522
Fast		
TB0 (cm)	10.2	9.04–11.4
TG (cm/week)	0.328	0.252–0.405
TS (1/week)	0.00362	−5.95E-04 to 0.00783
TP1[e]	1.20	0.719–1.67
TP2[e]	0.878	0.456–1.30
LDH effect on TB0[f]	0.868	0.752–0.984
LDH effect on TG[f]	0.771	0.473–1.07
Exposure (Cavg1) effect on TG[f]	−0.00342	−0.00690 to 6.69E-05
Random effects		
ω^2TB0	0.535 (0.731)	0.451–0.619
Fast		
ω^2TG	0.360 (0.600)	0.226–0.493
ω^2TS	4.07 (2.02)	0.690–7.45
No growth		
ω^2TS	0.385 (0.621)	0.135–0.636
ω^2TB_{ss}	1.38 (1.17)	0.857–1.90
Intermediate		
ω^2TG	0.203 (0.451)	−0.239 to 0.645
ω^2TS	3.21 (1.79)	0.999–5.41
ωTG (fast): ωTS (fast)	−1.06 (−0.878)	−1.62 to −0.506
ωTG (intermediate): ωTS (intermediate)	0.129 (0.159)	−0.622 to 0.879
Residual error		
Additive error (cm)	0.125	0.0927–0.158
Proportional error (–)	0.167	0.159–0.175

Cavg1, time-averaged concentration after the first dose; CI, confidence interval; LDH, lactate dehydrogenase; TB0, baseline tumor burden; TB_{ss}, steady-state tumor burden; TG, tumor growth; TGD, tumor growth dynamics; TS, tumor shrinkage.

[a]Parameters with fixed values (not estimated) are denoted with a superscript "f" after the names, with the fixed value given in the estimate column. [b]Random effects and residual error parameter names containing a colon denote correlated parameters. [c]Random effects and residual error parameter estimates are shown as variance (SD) for diagonal elements ($\omega_{i,i}$ or $\sigma_{i,i}$) and covariance (correlation) for off-diagonal elements ($\omega_{i,j}$ or $\sigma_{i,j}$). [d]Confidence intervals of random effects and residual error parameters are for variance or covariance. [e]TP1 and TP2 are the parameter estimates that determined the overall probability for each subpopulation, representing the approximate fraction of patients in the analysis data set in each subpopulation: mixture 1 (fast TG) and 2 (no-growth). The sum of overall probabilities was 1 and the overall probability of subpopulation 1, 2, and 3 are given by using the following equation:

Mixture 1: Fast TG: $\frac{TP1}{1+TP1+TP2}$

Mixture 2: no growth: $\frac{1}{1+TP1+TP2}$

Mixture 3: intermediate TG & TS: $\frac{TP2}{1+TP1+TP2}$

[f]The covariate effect on typical values (model estimated geometric mean) TG and TB0 are described by the following expression (BLDHU is the normalized baseline LDH with upper limit of normal):

$$TG_{Tv} = TG_{REF} \times (1 + \log(BLDHU) \times TG_{BLDHU}) \times (1 + Cavg1 \times TG_{Cavg1})$$

$$TB0_{Tv} = TB0_{REF} \times (1 + \log(BLDHU) \times TG_{BLDHU})$$

effect after taking into account all other covariate effects in OS-model 1. OS-model S2 did not further improve the model fit, as indicated by higher BIC value, suggesting that PRW8 provided sufficient information for predicting OS. The third sensitivity analysis was performed to assess whether subpopulation has an impact on OS after accounting for all other covariate effects in OS-model 1. The reduced BIC value in OS-model S3 relative to OS-model 1 suggested further improvement in the fit. The fourth sensitivity analysis was performed to assess whether two tumor response

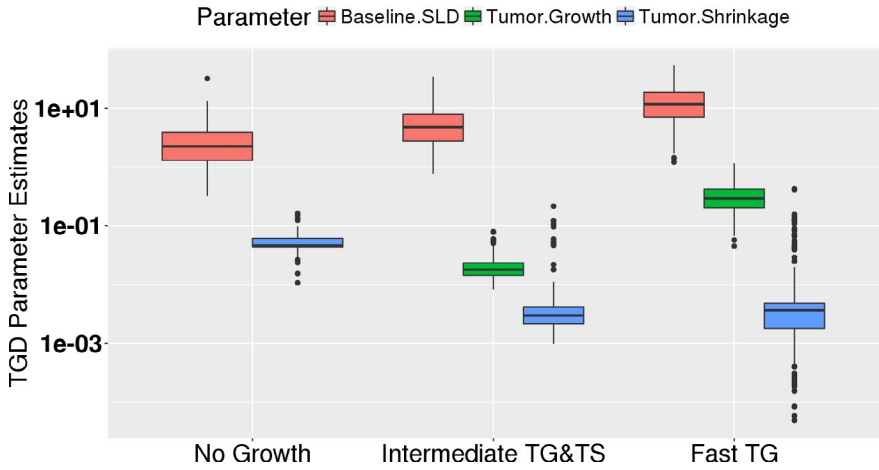

Figure 1 The distribution of tumor growth dynamics (TGDs) mixture model parameters by subpopulation: no-growth, intermediate tumor growth (TG) and tumor shrinkage (TS), and fast TG. The box plot shows the median and interquartile range of TGD parameter estimates in each group. SLD, sum of the longest diameter.

measures (PWR8 and CTB8, OS-model S4) might better predict OS relative to a single measure (PWR8, OS-model S3). The BIC was not further reduced, indicating that the addition of CTB8 did not meaningfully improve the goodness-of-fit of OS-model S3 (**Table S3**).

Model evaluations were conducted for OS-model 1 (final OS model) and sensitivity analysis models (OS-model S1 and OS-model S3). Model evaluation was performed by a VPC comparing the observed K-M curve with the model predicted median 90% PI K-M curve, obtained by simulation of events from the predicted individual survival curves. The evaluation was performed as an internal validation with the model-building data set from CA184-169. Model evaluation of OS-model 1 showed that the model-predicted median (90% PI) of OS was consistent with the observed K-M curve in both the 3 mg/kg and 10 mg/kg groups (**Figure 4a**), although the K-M curve of the observed data was at the upper edge of the 90% PI for the higher dose arm. The model-predicted median (90% PI) of OS for each dose arm categorized by TGD subpopulations was consistent with the observed K-M curve (**Figure 4b**). The K-M curves were in reasonable agreement with the CPH model predictions, although slightly below the prediction for the no-growth population at 10 mg/kg. Sensitivity analysis of OS-model S1 underpredicted results in the no-growth group and overpredicted in the fast TG group, indicating that a model using nonmixture PRW8 might not be able to describe OS in poor or improved OS populations (**Figure S3**). Interestingly, compared with OS-model 1, VPC in sensitivity analysis (OS-model S3) that included a TGD model subpopulation effect on OS showed a marked improvement in describing data in the no-growth subpopulation (**Figure S4**). This suggests that the association of PRW8 and OS in each subpopulation might not be the same. Results from model evaluation suggested that the OS model performance would be sensitive to a measure of tumor responses derived from different structure TGD models (mixture vs. nonmixture; **Figure S5**).

The predicted survival events from the VPC were used to estimate a predicted HR (median, 95% PI) for the treatment effect (for the overall population and by TGD-model subpopulations), as well for the HR with respect to the subpopulations. The predicted HR for the overall treatment effect was 0.92 (**Table S4**), whereas the treatment effect within a subpopulation was closer to 1 (0.96–0.98). Notably, the HR for the intermediate and no-growth subpopulations relative to the fast TG were 0.53 and 0.33, respectively (**Table S5)**. This suggests that the differences in the overall HR are due to differences in the percentage of patients in each subpopulation within each treatment group. Indeed, **Table S2** shows that the percentage of patients in the intermediate and no-growth subpopulations are higher in the 10 mg/kg dose group than in the 3 mg/kg dose group, which may in part explain the better OS in the 10 mg/kg dose group.

DISCUSSION

A major challenge in oncology drug development is the uncertain relationship between the gold standard end point of OS, which is commonly preferred for confirmatory phase III studies, and RECIST ORR, which is commonly used for go/no-go decisions in early clinical development. The situation has grown more complex in the era of immunotherapies, in which improvements in OS may be observed even without achieving a tumor response. Study CA184-169 is a case in point, where OS was better with ipilimumab 10 mg/kg compared with ipilimumab 3 mg/kg, even though the ORR and progression-free survival were similar.[22]

Intuitively, a continuous measure of the tumor response that also accounts for the dynamics of the response might be a better predictor of OS than ORR by RECIST criteria. The tumor response measures assessed as predictors of OS were limited to those that can be determined with data collected by week 8, to minimize the potential of a guarantee-time bias. In particular, this type of bias is less likely with a week 8 tumor measure, which is assessed at a fixed timepoint relatively soon after the start of treatment, compared with other measures, such as time-to-tumor growth. All measures of tumor response improved the fit of the OS model. PRW8 was

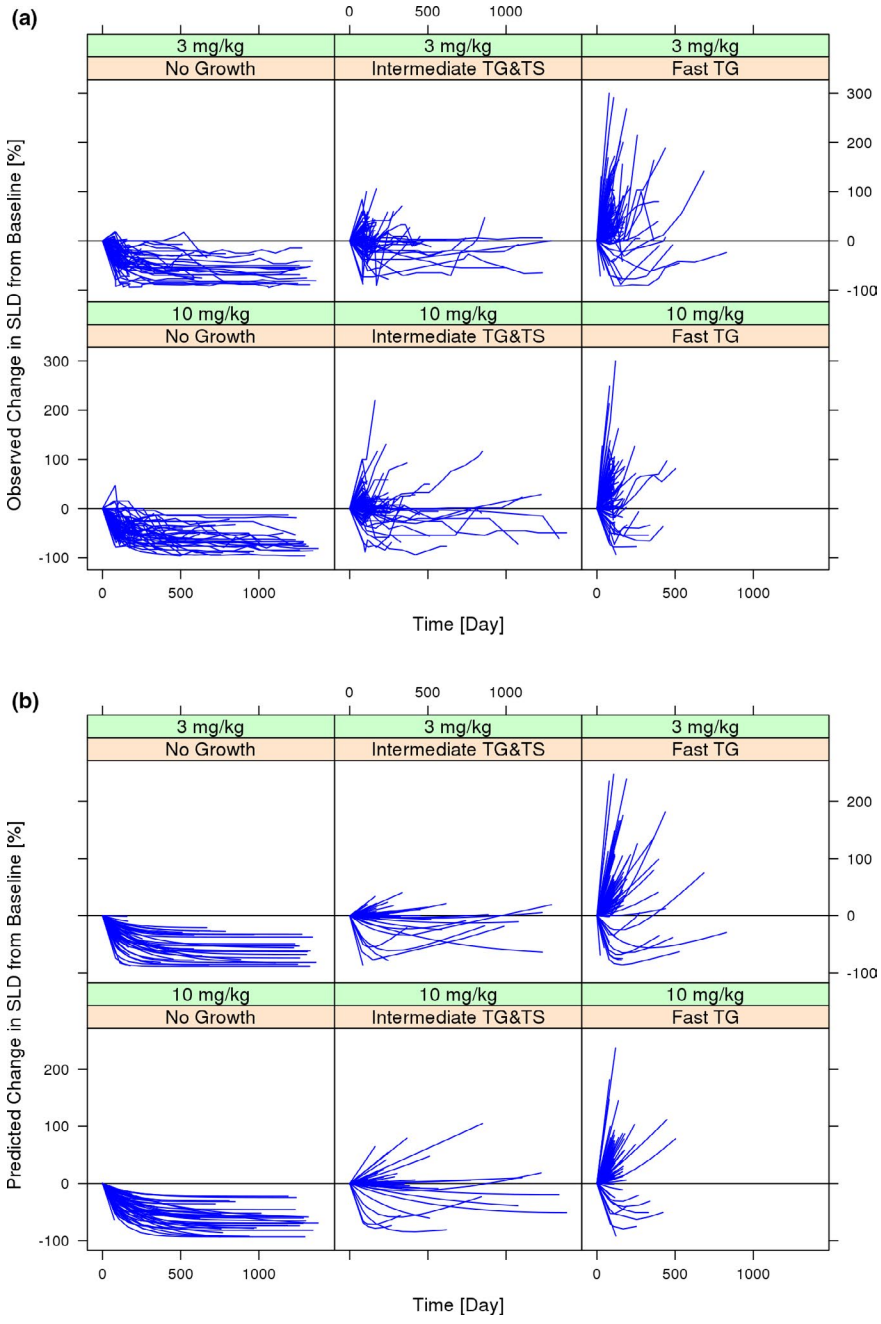

Figure 2 Individual time course of (**a**) observed and (**b**) predicted change in tumor size from baseline stratified by dose and subpopulation. SLD, sum of the longest diameter; TG, tumor growth; TS, tumor shrinkage.

a slightly better predictor of OS than CTB8, and both of these summary measures of tumor response were better than ORR and markedly better than either TG or TS. This may be due to TG and TS being less precisely determined than PRW8 or CTB8, whereby correlated values of TG and TS may result in similar values of PRW8 and CTB8. Furthermore, adding a treatment effect did not improve the model, indicating that the PRW8 is a sufficient statistic to explain the treatment effect. This suggests that PRW8 in early clinical trials holds promise for use in assessing the potential OS benefit of alternative treatments.

The OS analysis also highlighted the importance of TGD model selection. The three-mixture subpopulations TGD model provided a better description of the OS data (lower BIC) than the nonmixture model, likely because the mixture model was better able to describe the patterns of tumor response observed with immunotherapy, including steady-state TB. Distinct distributions of TGD parameters were identified, which reflects the three different patterns of tumor dynamics: no-growth, intermediate TG and TS, and fast TG subpopulations. Most previously reported TGD models were developed to describe response to chemotherapy or

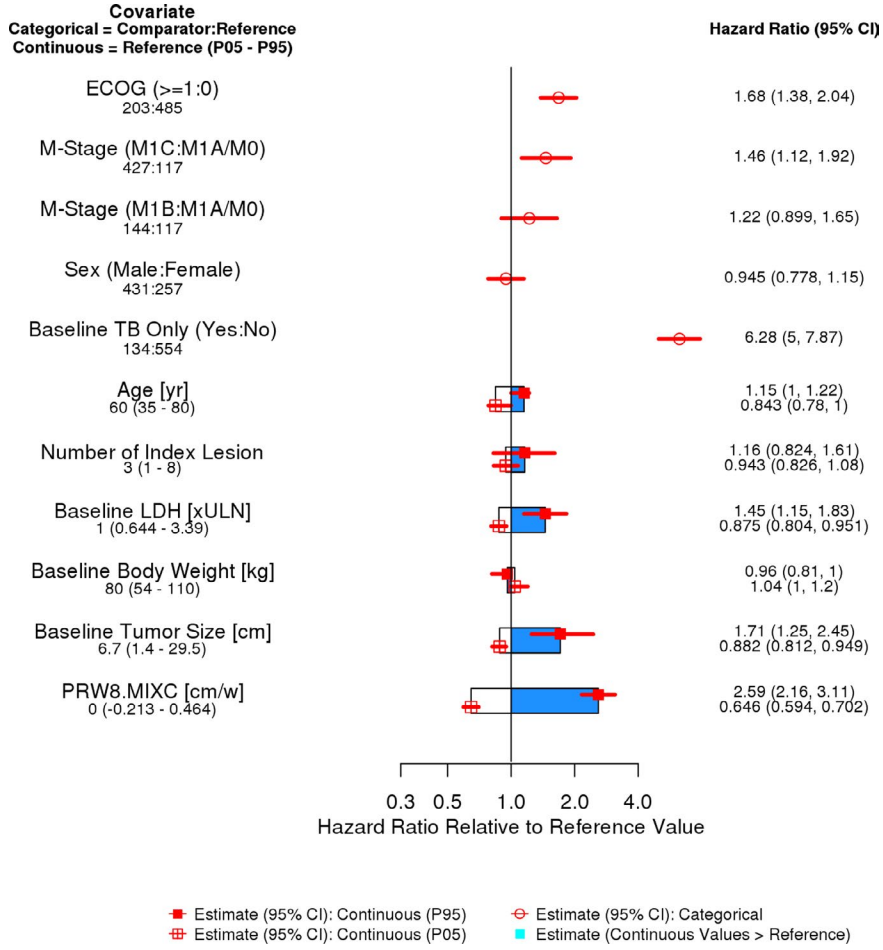

Figure 3 Effect of covariates on the hazard ratio of overall survival (OS-model 1). Missing postbaseline tumor burden (TB), patients only had baseline TB assessment and without post treatment TB assessment. CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; LDH, lactate dehydrogenase; PRW8.MIXC, progressive rate at week 8 from covariate mixture model; ULN, upper limit of normal.

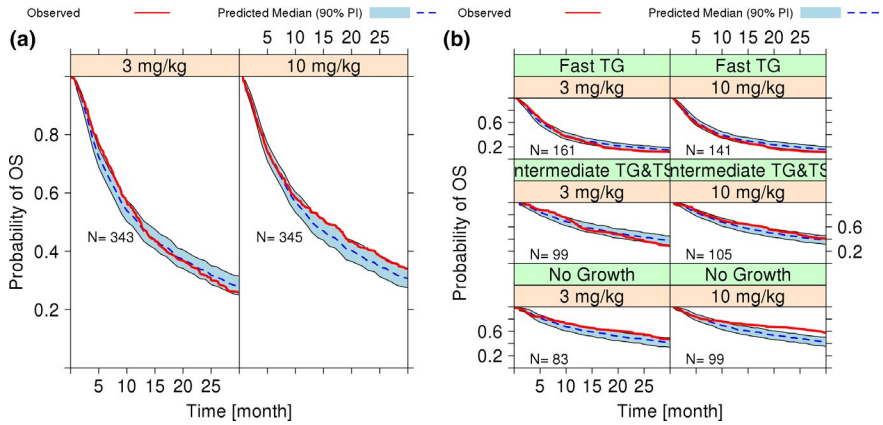

Figure 4 Model evaluation of overall survival (OS)-model 1 analysis stratified (**a**) by treatment and (**b**) by treatment and subpopulation. PI, prediction interval; TG, tumor growth; TS, tumor shrinkage.

targeted therapy, in which the response to therapy generally is transient. The structure of the previously published TGD models is such that the tumor is always changing in size (either only growing, or shrinking and then growing).[1,2] Therefore, the previously published nonmixture TGD models are not able to describe the pattern of steady-state TB observed with immunotherapy. Mixture models for immunotherapy have been proposed previously to describe the

delay in the response,[11] and have been explored to describe heterogeneity in the tumor response. However, the value of describing the heterogeneity in the response was not fully appreciated without the link to OS.[12]

The OS model provided reasonable description of observed data in both dose and subpopulation groups. Notably, the OS model using a nonmixture model-derived tumor metric was unable to provide a reasonable description of OS in patients who had poor or improved survival. Mixture modeling provided more degrees of freedom for TGD parameter estimations in each subpopulation, whereas the individual parameter estimation in a nonmixture model tended to shrink to 0 (**Figure S5**), leading to less discrimination of the effect of TGD on OS and, therefore, poor performance in OS model prediction (**Figure S3**). Moreover, as expected, interindividual variability of TGD parameters from the mixture model was generally smaller than that from the nonmixture model (**Figure S6**). The performance of the final model (OS-model 1) illustrated the importance of precise characterization of individual tumor dynamics.

Evaluation of exposure-response poses a challenge for the TGD model, and, in the current analysis, the effect of Cavg1 on TGD model parameters was not significant. However, there was an association between dose and mixture subpopulation, with the fraction of patients in the no-growth subpopulation higher with ipilimumab 10 mg/kg (29%) vs. 3 mg/kg (24%), indicating that more patients may achieve durable responses with higher doses of ipilimumab.

Because tumor dynamics are different among the three subpopulations, we also evaluated the TGD subpopulation as a predictor variable in the OS model (OS-model S3). The OS model that included subpopulations and PRW8 had the lowest BIC relative to all other tested base models with a single tumor metric. This suggests that, after accounting for the tumor metric effect (PRW8) on OS, the hazard of death was still different between subpopulations. Although utilization of PRW8 as predictor does violate the strict definition of guarantee-time bias as the value is not known at the start of treatment, the bias is likely minimal as the values are known at week8, which is early relative to the 2-year timespan for the death events.

Although the predicted HR of the treatment effect tends to underestimate the extent of the benefit associated with the higher ipilimumab dose, the 95% CI does include the HR of 0.84 determined from the observed data (**Table S4**). This finding indicated that, although the OS predicted using the TGD model for tumor response at week 8 was consistent with the observed data, the model was not as sensitive in discerning statistically significant differences in the treatment HR as the observed OS. Nonetheless, the predicted OS was markedly better for the no-growth and intermediate TG/TS subpopulations relative to the fast growth subpopulations (**Figure S4b**). The lack of sensitivity with the current TGD-OS model with respect to the HR between the treatment groups could be potentially addressed by utilizing longitudinal tumor response measures as a predictor of OS, instead of PRW8. There may, therefore, be value in a more complex OS model that incorporates time-varying measures of response that can incorporate additional features of the longitudinal tumor response without introducing guarantee-time bias.

Our results are consistent with previously published reports of the association between OS and change in TB at week 8 in several tumor types.[1,7,12] However, unlike previously reported analyses of TGD OS, our OS model also included an effect for missing postbaseline TB, which was a highly significant negative predictor of the risk of death (**Figure 3**; HR, 6.28; 95% CI, 5.00–7.87). The likely reason for this is that postbaseline tumor assessments may not be recorded for patients who progress rapidly; this is informative missing data (**Figure S2**). Therefore, it is important to include the effect of missing postbaseline tumor assessment, particularly when there is an imbalance in the percentage of patients with missing data in the treatments being compared.

In summary, the mixture modeling showed a better ability to describe individual tumor profiles and characterize different patterns of tumor dynamics. It also enabled robust estimation of exposure effects on tumor dynamics. The precise estimation of individual tumor dynamics is critical for OS prediction, and the TGD-OS model using derived measures of tumor response from a mixture model provided better descriptions of the OS data in each subpopulation, compared with results from a nonmixture model.

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

Figure S1. TGD mixture model evaluation. (**a**) Observed vs. predicted percentage change of tumor burden (TB) from baseline at first scan at week 12. (**b**) Box plot of observed vs. predicted percentage change of TB from baseline at first scan at week 12.
Figure S2. Kaplan-Meier curve of OS by BOR.
Figure S3. Sensitivity of OS analysis (OS-Model S1) using PRW8 derived from the non-mixture TGD model stratified by (**a**) treatment and (**b**) by treatment and subpopulation.
Figure S4. Sensitivity OS analysis (OS-Model S3) with effect of both subpopulation and PRW8 derived from the mixture TGD model stratified by (**a**) treatment and (**b**) by treatment and subpopulation.
Figure S5. Distribution of progression rate at week 8 (PRW8) derived from the mixture model and non-mixture model in each subpopulation.
Figure S6. Distribution of TGD parameters (TS and TG) obtained from the mixture model (TGD-Model 11) and non-mixture model (TGD-Model 4) in each subpopulation.
Table S1. TGD model development.
Table S2. Percentage of patients in each subpopulation by treatment arm.
Table S3. OS model development and sensitivity analysis.
Table S4. Predicted HR using OS-Model S3 between ipilimumab 10 mg/kg and 3 mg/kg doses (3 mg/kg as reference).
Table S5. Predicted HR using OS-Model S3 between subpopulations (Fast TG as reference).

Acknowledgments. The authors would like to thank Dave Clawson from Bristol-Myers Squibb for supporting the analysis data set. The authors would also like to thank Lora Hamuro and Curtis Johnston from Bristol-Myers Squibb and Johnathan French and Ramon Garcia from Metrum Research Group for their support and comments on the analysis. Editorial support was provided by Kakoli Parai, PhD, and Andrea Lockett at StemScientific, an Ashfield Company, funded by Bristol-Myers Squibb.

Funding. This study was funded by Bristol-Myers Squibb.

Conflict of Interest. Y.F., S.S., A.B., and A.R. are employees of Bristol-Myers Squibb. Y.F., A.B., A.R., and X.W. hold stock in Bristol-Myers Squibb.

Author Contributions. Y.F., X.W., S.S., A.B., and A.R. wrote the manuscript. Y.F., X.W., and A.R. designed the research. Y.F. performed the research. Y.F. analyzed the data.

Data Sharing. The Bristol-Myers Squibb policy on data sharing may be found at https://www.bms.com/researchers-and-partners/independent-research/data-sharing-request-process.html.

1. Wang, Y. *et al.* Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin. Pharmacol. Ther.* **86**, 167–174 (2009).
2. Stein, W.D. *et al.* Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. *Clin. Cancer Res.* **17**, 907–917 (2011).
3. 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).
4. Tham, L.S. *et al.* A pharmacodynamic model for the time course of tumor shrinkage by gemcitabine + carboplatin in non-small cell lung cancer patients. *Clin. Cancer Res.* **14**, 4213–4218 (2008).
5. Houk, B.E., Bello, C.L., Poland, B., Rosen, L.S., Demetri, G.D. & Motzer, R.J. Relationship between exposure to sunitinib and efficacy and tolerability endpoints in patients with cancer: results of a pharmacokinetic/pharmacodynamic meta-analysis. *Cancer Chemother. Pharmacol.* **66**, 357–371 (2010).
6. Bonate, L. Modeling tumor growth in oncology. In Pharmacokinetics in Drug Development: Advances and Applications (Bonate, P.L. and Howard, D.R., eds.) 1–19 (Springer, New York, NY, 2011).
7. Ribba, B. *et al.* A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e113 (2014).
8. Feng, Y., Agrawal, S., Gupta, M., Gupta, A.K., Masson, E. & Roy, A. Association between immune-checkpoint inhibitor induced tumor shrinkage and overall survival in advanced melanoma and NSCLC. *J. Clin. Oncol.* **32**, abstract 3053 (2014).
9. Suryawanshi, S., Godfrey, C.J., Roy, A., Simonsen, K., French, J. & Gupta, M. Leveraging longitudinal tumor data for prediction of overall survival from I-O therapy: a proof of concept with quantitative models and external validation. *J. Pharmacokinet. Pharmacodyn.* **42**, S11–S107 (2015).
10. Claret, L. *et al.* A model of overall survival predicts treatment outcomes with atezolizumab vs chemotherapy in non-small cell lung cancer based on early tumor kinetics. *Clin. Cancer Res.* **24**, 3292–3298 (2018).
11. 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).
12. Chatterjee, M.S. *et al.* Population pharmacokinetic/pharmacodynamic modeling of tumor size dynamics in pembrolizumab-treated advanced melanoma. *CPT Pharmacometrics Syst. Pharmacol.* **6**, 29–39 (2017).
13. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185**, 71–110 (1894).
14. Feng, Y. *et al.* Population pharmacokinetic analysis for risperidone using highly sparse sampling measurements from the CATIE study. *Br. J. Clin. Pharmacol.* **66**, 629–639 (2008).
15. Choi, S., Zhu, L. & Huang, X. Semiparametric accelerated failure time cure rate mixture models with competing risks. *Stat. Med.* **37**, 48–59 (2018).
16. Sondak, V.K., Smalley, K.S., Kudchadkar, R., Grippon, S. & Kirkpatrick, P. Ipilimumab. *Nat. Rev. Drug Discov.* **10**, 411–412 (2011).
17. Hoos, A. *et al.* Development of ipilimumab: contribution to a new paradigm for cancer immunotherapy. *Semin. Oncol.* **37**, 533–546 (2010).
18. Hodi, F.S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–713 (2010).
19. Robert, C. *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).
20. Wolchok, J.D. *et al.* Ipilimumab monotherapy in patients with pretreated advanced melanoma: a randomised, double-blind, multicentre, phase 2, dose-ranging study. *Lancet Oncol.* **11**, 155–164 (2011).
21. Ascierto, P.A. *et al.* Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol.* **18**, 611–622 (2017).
22. Vezina, H.E., Wang, S., Statkevich, P., Bello, A. & Roy, A. Exposure-response analysis to support benefit-risk assessment of ipilimumab 3 vs 10 mg/kg in patients with advanced melanoma. *J. Pharmacokinet. Pharmacodyn.* **45**, S3–S134 (2018).