The AAPS Journal (2019) 21: 22
DOI: 10.1208/s12248-018-0290-x

Research Article

Predicting Overall Survival and Progression-Free Survival Using Tumor Dynamics in Advanced Breast Cancer Patients

Hyeong-Seok Lim,[1] Wan Sun,[2] Kourosh Parivar,[2] and Diane Wang[2,3]

Received 3 September 2018; accepted 17 December 2018; published online 30 January 2019

ABSTRACT. Prediction of survival endpoints, *e.g.*, overall survival (OS) and progression-free survival (PFS), based on early observations, *i.e.*, tumor size, may facilitate early decision making in oncology drug development. In this paper, using data from six randomized trials for first- or second-line advanced breast cancer (ABC) treatments with various mechanisms of action, tumor size change from baseline at different observation time points was evaluated as a predictor for survival endpoints using different modeling approaches. The aim is to establish a predictive model where tumor size change from baseline can be used as a treatment independent predictive marker for PFS and OS in first- and second-line ABC. The results showed that tumor size change at single time point (TSP) or up to certain time points as a time-varying covariate (TSTVC) were significant predictors for OS and PFS in the survival models along with other covariates identified for each line of treatment. TSP and TSTVC models performed similarly for first-line treatments; TSTVC performed significantly better for second-line treatments. Eight weeks was selected as the recommended early evaluation time of tumor size change to predict OS and PFS in both first- and second-line treatment, while better prediction can be achieved for first-line OS by using 16 weeks tumor size change. The result of this study is treatment independent and can be used to predict the outcome of the clinical trials using early readout of tumor size change for the classes of drugs that have been evaluated in this study.

KEY WORDS: breast cancer; OS; PFS; time-varying; tumor size.

INTRODUCTION

Significant effort has been made toward identifying predictors of oncology efficacy endpoints that are used for regulatory approval, such as overall survival (OS) and progression-free survival (PFS), to reduce attrition in late-phase drug development. Models have been developed to evaluate the relationships between early response measurements and efficacy endpoints for a variety of tumor types (1–8). These analyses have shown that tumor size change from baseline as an early response measurement and other risk factors can be predictive markers for OS and PFS. The use of these models to inform "go/no-go" decisions and trial design is consistent with recommendations of the United States Food and Drug Administration and the European Medicines Agency for drug development (9,10).

Compared to other more aggressive tumor types, breast cancer patients have a relatively longer PFS and OS given the disease characteristics. As a result, it requires a longer period of time to evaluate the efficacy outcome of test therapeutic agents in clinical trials for breast cancer; thus, a reliable prediction for the efficacy outcomes using identified prognostic factors becomes even more crucial and useful for trial design and drug development. Tumor size change from baseline has been used to predict the clinical outcomes of a phase 3 trial using a model built based on Phase 2 data with the same patient population and treatment (2), and simulate the clinical outcome of a future trial at a different dose level using a model built based on data from two completed trials with the same patient population and treatment (6). However, modeling work including data from multiple drugs, especially those with different mechanisms of action for breast cancer treatment to evaluate the prediction value of

Hyeong-Seok Lim participated in this study during an international fellowship at Pfizer.

Wan Sun was employed at Pfizer Inc at the time of study conduct and manuscript development.

Electronic supplementary material The online version of this article (https://doi.org/10.1208/s12248-018-0290-x) contains supplementary material, which is available to authorized users.

[1] Department of Clinical Pharmacology and Therapeutics, Asan Medical Center, University of Ulsan, Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, Republic of Korea.
[2] Clinical Pharmacology, Global Product Development, Pfizer, 10555 Science Center Dr., San Diego, California, 92121, USA.
[3] To whom correspondence should be addressed. (e–mail: Diane.Wang@pfizer.com)

tumor size change from baseline, has not been reported. In addition, systematic investigations on the predictive performance of different modeling approaches of using tumor size change and different observation times for tumor size change as a predictor of efficacy outcome are currently not available.

In the present study, models describing tumor size change over time and its relationship to survival endpoints were developed using data from six randomized clinical trials in patients with advanced breast cancer (ABC), including three first-line trials and three second-line trials. These trials included various anti-cancer agents with different mechanisms of action including cytotoxic chemotherapies, anti-angiogenesis, and a targeted therapy. The aim was to establish a predictive model where tumor size change from baseline could be used as a treatment independent predictor for PFS and OS outcome in first- and second-line treatment of ABC. In this study, we evaluated (1) different ways of using tumor size change data in the survival model for each line of treatment, namely tumor size change at single time point as a covariate (TSP) versus longitudinal tumor size change as a time-varying covariate (TSTVC) in the survival model; (2) the timing of the tumor assessment that will provide a reasonable prediction of PFS and OS as an early treatment response; and (3) different imputation methods for tumor size post last tumor evaluation when tumor size change is used as a time-varying covariate. The result of this study, if it is treatment independent, can be used to predict the outcome of the clinical trials using early readout of tumor size change for the classes of drugs that have been evaluated in this study.

METHODS

The analysis database included six randomized clinical trials for ABC treatments, with 1133 patients from 3 first-line treatment trials (one Phase 2 and two Phase 3 trials) and 852 patients from 3 second-line treatment trials (two Phase 2 and one Phase 3 trials). As shown in Table I, these studies evaluated various agents with different mechanisms of action, *i.e.*, cytotoxic chemotherapies (docetaxel, paclitaxel, bevacizumab, capecitabine), anti-angiogenesis (axitinib and sunitinib), and targeted therapy (neratinib and lapatinib), given as first-line or second-line therapies.

First, baseline prognostic factors were screened in a parametric survival model to identify the factors that had significant impact on survival outcomes PFS and OS, separately for first-line and second-line treatments. The models included these significant prognostic factors were referred as semi-final survival models. Then, a tumor size model was developed to describe the tumor size change over time for patients with both available baseline and post-treatment tumor size data. Subsequently, the model-predicted tumor size change was added to the semi-final models to link tumor response to PFS/OS time using different approaches.

The goodness-of-fit of the models to observed data was evaluated using statistical methods, diagnostic plots, and visual predictive check (VPC). The Wald statistics test was used for model evaluations, and the likelihood ratio test was used to compare hierarchical models. All model development and parameter estimations were performed using NONMEM 7.2 (ICON Development Solutions, Dublin, Ireland). Dataset organization and plotting were performed using R software, version 3.11.1 (R Foundation for Statistical Computing, Vienna, Austria). Data were expressed as the count, percentage, mean ± standard deviation (SD), or median (range), as appropriate.

All the study protocols were reviewed and approved by the Institutional Review Board at every participating hospital, and all participating patients provided written informed consent before screening. All these studies were conducted in compliance with the Declaration of Helsinki and the

Table I. Clinical Trials Included in the Analysis

Study (phase)	Main inclusion criteria	Treatment	Patients, n/N^a	Trial registration[b]
First-line			1122/1133	
Study 1 (Phase 2)	Metastatic BC	Axitinib + docetaxel versus docetaxel + placebo	162/171	NCT00076024 (11)
Study 2 (Phase 3)	HER2-negative BC with evidence of unrespectable locally recurrent or metastatic breast	Sunitinib + docetaxel versus docetaxel	538/538	NCT00393939 (12)
Study 3 (Phase 3)	Advanced BC	Sunitinib + paclitaxel versus bevacizumab + paclitaxel	422/424	NCT00373256 (13)
Second-line			829/852	
Study 4 (Phase 2)	Triple receptor-negative, locally recurrent, or metastatic BC	Sunitinib versus standard chemotherapy	191/204	NCT00246571 (14)
Study 5 (Phase 3)	Locally advanced or metastatic disease	Sunitinib + capecitabine versus capecitabine	407/407	NCT00435409 (15)
Study 6 (Phase 2)	Stage IIIB, IIIC, or IV HER2-positive breast cancer	Neratinib versus lapatinib + capecitabine	231/241	NCT00777101 (16)
Total			1951/1985	

BC breast cancer, *HER2* human epidermal growth factor receptor
[a] Number of patients included in the survival analysis/total number of patients with measurable disease
[b] ClinicalTrials.gov (https://clinicaltrials.gov/ct2/home)

International Conference on Harmonisation Good Clinical Practice Guidelines.

Evaluation of Prognostic Factors in Parametric Survival Models

The OS and PFS parametric models were developed separately for the first- and second-line treatments given the distingue survival curves in these different patient populations. The accelerated failure time (AFT) model was adopted to determine the relationship between covariates and the clinical outcomes (OS and PFS). Different from the widely used proportional hazards model that assumes a covariate affects the hazard, the AFT model assumes that the covariates affect the predicted survival and provides information in the form of a percentage difference in the expected time to an event according to the values of covariates (17,18). Furthermore, unlike proportional hazards models, the regression parameter estimates from AFT models are robust for omitted covariates and are less affected by the choice of probability distribution (19–21).

Exponential, Weibull, and log-logistic models were tested to describe OS and PFS data for first-line or second-line treatments. Likelihood ratio tests and diagnostic plots were applied for selection of an appropriate survival function. With the selected survival function, a univariate analysis was first performed to screen the potential covariates individually for OS and PFS in first- and second-line treatments using statistical significance level of $P < 0.01$, including demographic factors age, height, race, ethnicity, baseline body weight, and baseline body surface area; baseline laboratory test variables ALT, AST, bilirubin, creatinine clearance, serum creatinine, and lactate dehydrogenase; baseline disease characteristics ECOG performance status (0 versus ≥ 1), stage of breast cancer at the time of screening, ER status (negative versus positive), PR status (negative versus positive), HER2 status (negative versus positive), baseline sum of the longest diameters (BSLD), breast cancer subtype (ductal, lobular carcinoma, or adenocarcinoma); as well as other factors, *e.g.*, smoking status, study effect, treatment arm, and concomitant medications. Next, a stepwise addition process with statistical criteria of $P < 0.01$ was started to add each significant factor identified in the univariate analysis, and then a backward elimination process was followed with statistical criteria of $P < 0.001$. The multivariate models were obtained from the last stage of the elimination process and were considered as semi-final survival models. Missing covariate values were imputed only if the percentages of missing values were low for the covariates involved in the analysis (< 15%): 1) for categorical covariates, missing values were imputed to the most common category of non-missing covariate data, and (2) for continuous covariates, missing values were imputed as the median of the non-missing values.

Tumor Size Model

Tumor size was defined as the sum of the longest diameters (SLD) of measurable target lesions by computed tomography at pre-treatment and post-treatment time points. Since not all patients had available tumor size data at the same post-treatment time points due to differences in trial design across studies, missed or delayed tumor assessments for individual patients, and missing values in assessment time, tumor size models were developed to provide predicted tumor sizes that were used to correlate with efficacy endpoints PFS and OS in survival analysis. The main diagnostic criterion was the match between individual predicted values and observed data.

Two types of tumor size models were developed, a linear tumor growth and exponential decay model (1) (linear model) and a Gompertz growth and exponential decay model (22) (Gompertz model), shown as Eqs. 1 and 2, respectively.

$$\mathrm{IPRED} = \mathrm{SLGR} \times \mathrm{TIME} + \mathrm{BSLD} \times \mathrm{EXP}(-\mathrm{KIL} \times \mathrm{TIME}) \tag{1}$$

$$\mathrm{IPRED} = \mathrm{TMS} \times \mathrm{EXP}(-\mathrm{DPT} \times \mathrm{EXP}(-\mathrm{GR} \times \mathrm{TIME})) + \mathrm{BSLD} \times \mathrm{EXP}(-\mathrm{KIL} \times \mathrm{TIME}) \tag{2}$$

where IPRED is predicted tumor size; BSLD is the observed baseline SLD; SLGR and GR are growth rate in the linear growth model and Gompertz growth model, respectively; KIL is the exponential tumor shrinkage rate constant; TMS is maximum tumor size achievable; and DPT is the difference between baseline tumor size and maximum tumor size in natural log scale. Inter-patient variabilities of fixed effect parameters were also taken into account in the form of log normal error model.

Final Survival Models

Model predicted tumor size change was added to the semi-final survival models to correlate with PFS or OS using two different modeling approaches. One approach is to use tumor size change at single time point as a covariate (TSP); one approach is to use tumor size change up to certain evaluated time points as a time-varying covariate (TSTVC). In the TSP approach, the predicted tumor size proportional changes (%) from baseline at single time points were incorporated in the semi-final survival model. In the TSTVC approach, the time course of tumor size proportional changes predicted from the beginning of the treatment up to certain evaluated time point was incorporated into the survival models as a time-varying covariate. Two different ways of incorporation were tested: one was to use the predicted tumor size changes for each patient until after the occurrence of event, and the other way was to use the predicted tumor size changes up to the selected evaluation time, such as 8 weeks for second-line PFS and OS, and then use the prediction at the selected evaluation time (8 weeks for second-line PFS and OS) for all following time points until after the occurrence of the event, *i.e.*, last prediction carried forward (LPCF).

To search for the earliest time when tumor size change can provide a reasonable prediction of OS and PFS, the performance of different observation durations in predicting OS and PFS was evaluated for each approach. The durations of tumor size change observation evaluated are every 2 weeks from 8 to 40 weeks post-treatment for OS and from 8 to

24 weeks post-treatment for PFS. In order to do that, multiple partial tumor size datasets were created by excluding the data after each evaluated time point to mimic the actual case scenario, *i.e.*, study is still ongoing and only short-term tumor size data are available from interim analysis. For instance, when evaluating the time point of 16 weeks post-treatment, a new tumor size dataset excluding data collected after 16 weeks post-treatment was to be created. Then, based on each of these partial tumor size datasets, individual tumor model parameters were obtained as empirical Bayesian estimates, where the final tumor model parameters estimated based on all available tumor size data were used as prior information. Subsequently, these individual tumor size model parameters were used to predict tumor size data for each patient, which was to be included in the survival model as a covariate.

The model predictive performance was evaluated by VPC. For VPC, 100 replicates of survival data were simulated using uniform distribution for each scenario. Two subgroups were created according to the identified prognostic factors or tumor size proportional change at each evaluated time point, *e.g.*, patients with tumor size change at 16 weeks equal or smaller than versus larger than the overall median tumor size change at 16 weeks in the entire population. Median survival curves and 95% CIs were then constructed across the 100 simulations for each subgroup and were compared with the observed survival curves stratified by subgroup in the Kaplan-Meier plots.

RESULTS

Patients and Data

The demographics and baseline disease characteristics of patients are summarized in Table II. Majority of patients were female and White with ECOG performance status as 0 or 1. In first-line treatment trials, majority of patients had ER+ ABC (72.9 versus 25.4%), while in second-line treatment trials, less than half of patients had ER+ ABC (44.1 versus 54.6%). In second-line trials, approximately 65.6% of patients were HER2-negative (versus 28.2% for HER2-positive) and 64.3% PR-negative (versus 32.4% for PR-positive).

Semi-Final Survival Model

Among the tested survival functions, Weibull was superior to exponential in all cases based on the log-likelihood ratio test. The objective function value (OFV) of log-logistic was lower than that of Weibull for OS (1717.785 versus 1728.959) and PFS (801.543 versus 993.707) with first-line therapy and for PFS (459.077 versus 660.077), but not OS (1703.204 versus 1691.371), with second-line therapy; however, statistical significance could not be tested. The predictive performance for OS and PFS was similar between the log-logistic and Weibull models from visual predictive check plots (data not shown). Thus, the log-logistic model was selected for the analyses of both OS and PFS.

As shown in Fig. 1, multiple significant covariates were identified for each endpoint in each line of treatment and were included in the semi-final model before testing tumor size change as a predictor. These semi-final models described the observed survival curve for each subgroup divided based on identified covariates reasonably well, which supports the significant roles of these covariates in predicting the survival outcomes. The effect of study and type of anti-cancer treatments patients received were also tested and shown not to be significant covariates after the above significant covariates were included.

Tumor Size Model

A total of 9494 tumor measurements from 1985 patients enrolled in the 6 trials were used to develop the tumor size model. While both linear and Gompertz models had reasonable performance and model parameters were estimated with acceptable precision, Gompertz model provided better prediction to observed longitudinal tumor size data with much lower OFV (data not shown). Therefore, predicted tumor size changes based on Gompertz model were to be used as a predictor in survival models. The goodness of fit plots for final tumor size model are shown in Fig. S1. In addition, the individual fitting plot for randomly selected patients based on partial dataset is also provided in this figure to demonstrate the reasonable model performance based on very limited tumor observations.

Inclusion of Tumor Size Change as a Covariate in the Survival Models

After tumor size change was added into the models, significant decreases in OFVs from semi-final models were observed in both first-line and second-line treatments. Figure 2 shows the OFVs of the OS and PFS models incorporating model-predicted tumor size change from baseline at or up to different observation time points as a covariate. In each of these OFV plots, three scenarios with different modeling approach were presented, including TSP as covariate, TSTVC with model-predicted tumor size change beyond the selected evaluation time as covariate, and TSTVC with LPCF beyond the selected evaluation time as covariate.

1. Comparison of approaches for tumor size change beyond evaluation time as a predictor for TSTVC modeling approach

When using TSTVC, LPCF approach performed better than using the predictions beyond the tumor size evaluation times for all four scenarios demonstrated by consistently lower OFVs (Fig. 2). Therefore, TSTVC with LPCF approach was used in final models to further evaluate observation timing of tumor size change.

2. Comparison of modeling approaches (TSP versus TSTVC)

As shown in Fig. 2, the OFVs from models with TSP are either comparable (OS) or lower than those with TSTVC (PFS) for first-line treatment, while the OFVs from models with TSTVC are significantly lower than TSP for second-line OS and PFS. The comparison between these two modeling approaches was also performed through VPC plots. In this analysis, patients were divided into two subgroups according to tumor size change at 16 weeks, and model-predicted median OS and PFS curves (with 95% prediction intervals)

Table II. Patient Demographics and Baseline Disease Characteristics

Characteristic	First-line treatment ($n = 1122$)	Second-line treatment ($n = 829$)
Female sex, *n* (%)	1121 (99.9)	825 (99.5)
Race/ethnicity, *n* (%)		
White	980 (87.3)	681 (82.1)
Black	55 (4.9)	28 (3.4)
Asian	46 (4.1)	88 (10.6)
Other	41 (3.7)	24 (2.9)
Missing	0 (0)	8 (1.0)
Median age (range), years	55.0 (28.0–92.0)	53.0 (27.0–81.0)
Median weight (range), kg	69.7 (39.0–154.5)	68.5 (37–126.3)
Missing, *n* (%)	0 (0)	127 (15.3)
Median body surface area (range), m^2	1.75 (1.26–2.54)	1.73 (1.24–2.39)
Missing, *n* (%)	2 (0.2)	7 (0.8)
Median baseline tumor size[a] (range), cm	6.8 (1.0–65.7)	6.2 (1.0–56.2)
Cancer stage at screening, *n* (%)		
IIIB	0 (0)	8 (1.0)
IIIC	0 (0)	4 (0.5)
IV	124 (11.1)	219 (26.4)
Recurrent	576 (51.3)	598 (72.1)
Missing	422 (37.6)	0 (0)
ECOG performance status, *n* (%)		
0	598 (53.3)	508 (61.3)
1	475 (42.3)	309 (37.3)
≥ 2	23 (2.0)	11 (1.3)
Missing	26 (2.3)	1 (0.1)
ER status, *n* (%)		
ER^+	818 (72.9)	366 (44.1)
ER^-	285 (25.4)	453 (54.6)
Missing	19 (1.7)	10 (1.2)
PR status, *n* (%)		
PR^+	560 (49.9)	269 (32.4)
PR^-	372 (33.2)	533 (64.3)
Missing	190 (16.9)[†]	27 (3.3)
HER2 status, *n* (%)		
$HER2^+$	12 (1.1)	234 (28.2)
$HER2^-$	936 (83.4)	544 (65.6)
Missing	174 (15.5)[b]	51 (6.2)

ECOG Eastern Cooperative Oncology Group, *ER* estrogen receptor, *HER2* human epidermal growth factor receptor 2, *PR* progesterone receptor

[a] Sum of the longest diameters of measurable tumor lesions by pre-treatment computed tomography scans

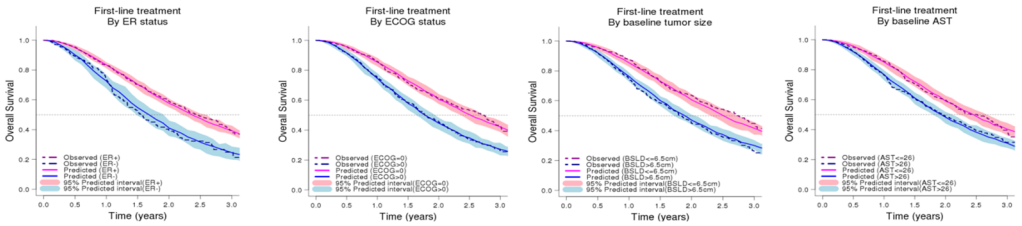
[b] Due to relatively high percentage of missing values, PR and HER2 status could not be tested as covariates for first-line treatment

for each subgroup based on 8- and 16-week evaluation times were compared with the observed Kaplan-Meier OS and PFS curves stratified by tumor size change at 16 weeks. In Fig. 3A, the TSP and TSTVC with LPCF models performed similarly in predicting OS for first-line treatment for both 8- and 16-week analyses, consistent with similar OFVs from TSP and TSTVC with LPCF models for first-line OS in Fig. 3A. Similar performance was also observed in predicting PFS for first-line treatment (Fig. 3C) despite lower OFVs from TSP models relative to those from TSTVC with LPCF models in Fig. 3. Consistent with comparison of OFVs in Fig. 3, OS and PFS predictions were significantly improved using the TSTVC compared to TSP models for second-line treatment (Fig. 3B, D). Based on this observation, TSP approach was used in the final models for first-line treatment and TSTVC approach was used in the final models for second-line treatment.

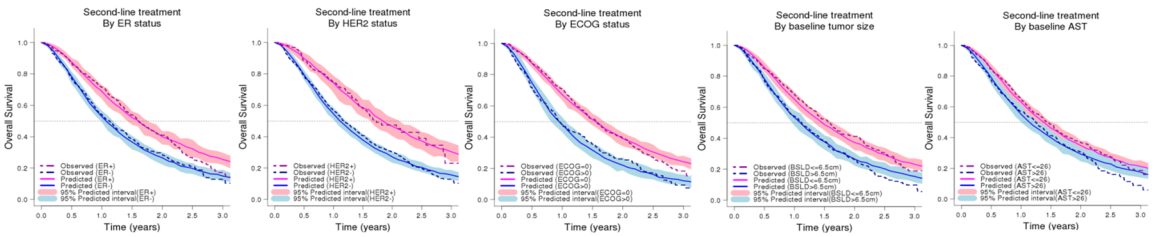
3. Comparison of the timing of tumor size evaluation as a predictor

Extensive evaluations were performed to determine the optimal evaluation time of tumor size change used for PFS and OS prediction. While there is a general trend of decreasing OFVs with longer follow-up of tumor size change, it plateaus beyond certain evaluation times for OS (26 weeks for first-line OS using TSP and 12 weeks for second-line OS using TSTVC) (Fig. 3). Using VPC plots, the effect of tumor evaluation time on OS and PFS prediction was further evaluated by comparing the predictive performance of evaluation time at 8, 16, and 24 weeks (data not shown) (Fig. 3). These time points were selected as possible earlier evaluation time points that still have predictive value and also provide reasonable OS and PFS predictions. For first-line OS, it is observed tumor size change at or up to 8 weeks provided

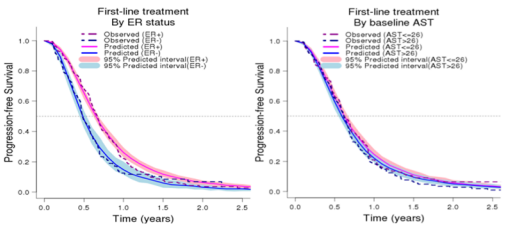
a ER status, ECOG status, baseline tumor size and baseline AST included in semi-final model for OS in the 1st-line treatment



b ER status, HER2 status, ECOG status, baseline tumor size and baseline AST included in semi-final model for OS in the 2nd-line treatment



c ER status and baseline AST included in semi-final model for PFS in the 1st-line treatment



d ER and HER2 status, ECOG status, and baseline AST included in semi-final model for PFS in the 2nd-line treatment

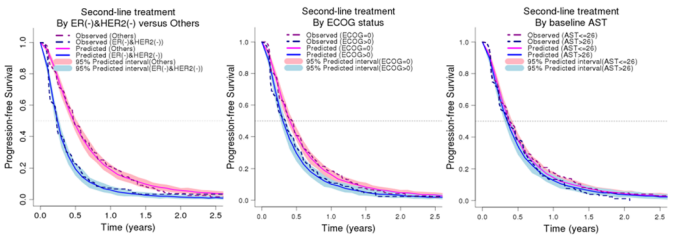


Fig. 1. Visual predictive check plots for OS and PFS semi-final models (prior to incorporating model-predicted tumor size change as a covariate). Comparison between observed and predicted survival curves for OS in the first-line treatment (**a**): subgroups with ER (+) versus ER (−), ECOG = 0 versus ECOG > 0, and BSLD > median value versus BSLD ≤ median value, and baseline AST > median value versus baseline AST ≤ median value; for OS in the second-line treatment (**b**): subgroups with ER (+) versus ER (−), HER2 (+) versus HER2 (−), ECOG = 0 versus ECOG > 0, and BSLD > median value versus BSLD ≤ median value, and baseline AST > median value versus baseline AST ≤ median value; for PFS in the first-line treatment (**c**): subgroups with ER (+) versus ER (−) and baseline AST > median value versus baseline AST ≤ median value; for PFS in the second-line treatment (**d**): subgroups with ER (−) plus HER2 (−) versus others, ECOG = 0 versus ECOG > 0, and baseline AST > median value versus baseline AST ≤ median value. The solid lines and shaded areas represent the median predicted survival curves and 95% confidence intervals, respectively. Broken lines represent the Kaplan-Meier survival curves based on observed data. BSLD, baseline sum of the longest diameter of the tumor; ECOG, Eastern Cooperative Oncology Group; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; AST, aspartate aminotransferase; KM, Kaplan-Meier; OS, overall survival; PFS, progression-free survival

reasonable prediction while 16-week evaluation time had better performance, consistent with the lower OFVs at 16 weeks (Fig. 3). However, there was no further improvement using 24-week evaluation time (data not shown) despite further decrease in OFV observed in Fig. 3. For first-line PFS, although OFV was lower at 16 weeks relative to those at 8 weeks (Fig. 3), using tumor size changes predicted at or up to 16 weeks did not yield better predictions for PFS compared to 8 weeks from VPC plots (Fig. 3C). For second-line treatments, the models with up to 16 weeks tumor size had lower OFVs but the improvement in both OS and PFS predictions was minimum compared to those with up to 8 weeks. Therefore, based on the predictive performance and also considering the value of early prediction for decision making in drug development, 8 weeks was selected as the recommended evaluation time of tumor size change to predict OS and PFS for both first-line and second-line treatments although 16 weeks can be used to achieve even better predictions for OS in first-line treatment.

Final Survival Model

The final survival model parameter estimates are shown in Table III. These parameters suggest that ER-negative status (versus positive), ECOG score ≥ 1 (versus 0), longer BSLD, and higher baseline AST are associated with shorter OS for first- and second-line treatments; HER2-negative status (versus positive) is also associated with shorter OS for second-line treatment. In first-line treatment, ER-negative status (versus positive) and higher baseline AST are associated with shorter PFS; in second-line treatment, ER-negative and HER2-negative status (versus others), ECOG score ≥ 1 (versus 0), and higher baseline AST are associated with shorter PFS. In all cases, greater tumor size shrinkage is associated with longer PFS/OS.

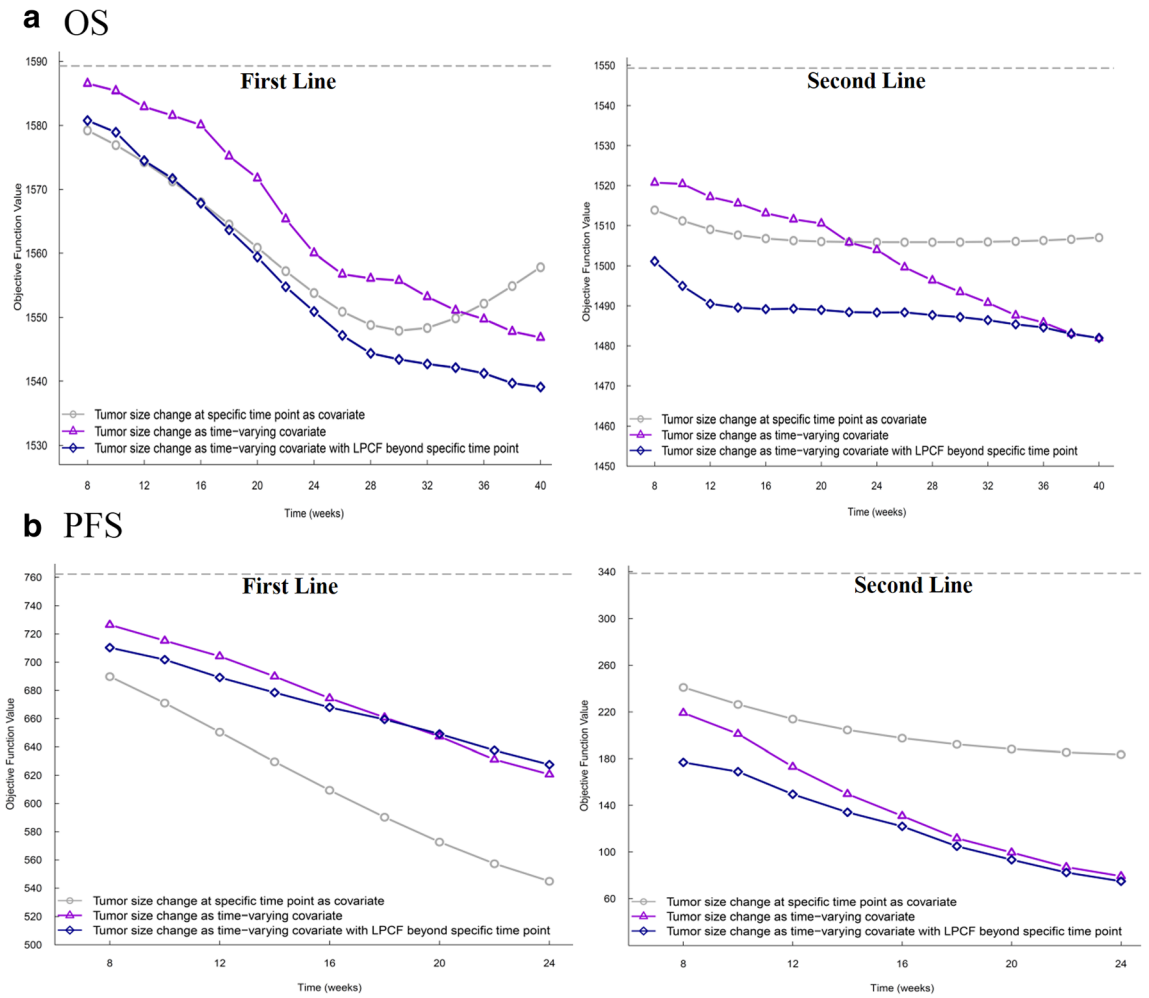


Fig. 2. Objective function value for the OS (**a**) and PFS (**b**) models at different observation time points using different modeling approaches. Dotted line in each figure represents the objective function value for each semi-final model. LPCF, last prediction carried forward; OS, overall survival; PFS, progression-free survival

Survival Model Evaluation

Evaluation of the final survival models was conducted through internal validation. The models were used to predict the survival curves of each treatment in each study (Fig. 4**)**. It can be seen that the model simulations provide reasonably good predictions of the observed survival curves for all individual treatment arms, confirming that there is no treatment or study effect.

DISCUSSION

The purpose of this study was to develop models that could predict OS and PFS based on early tumor response data in patients with advanced breast cancer receiving various types of anti-cancer treatments. The tumor size models were first developed to describe longitudinal tumor size change. As described in the "RESULTS" section, Gompertz model predictions were used in the survival analysis because of better predictive performance based on the current analysis data. In other cases where only sparse tumor size data are available, linear model may be more appropriate with fewer parameters and simpler model structure.

Based on the covariate analyses, ER/HER2 status, ECOG performance status, BSLD, and baseline AST were identified as significant covariates for OS or PFS in the first-line or second-line treatment settings as shown in Fig. 4 among all the potential covariates tested. ER status, ECOG, and baseline tumor size have consistently been recognized as prognostic factors in patients with breast cancer from other studies (23–27). Due to the high percentage of missing values (>15%), PR and HER2 status could not be tested for the first-line treatment.

Consistent with other published models used to predict efficacy endpoints in cancer patients (1–8,28–30), tumor size change from baseline was identified as a significant covariate to predict OS and PFS in the survival models. As it has been done in the above-published models, tumor size change at a single time point (TSP) was initially used as a tumor regression measure to correlate with OS and PFS. The modeling effort started with one of the Phase 2 trials (first-line Study 1) for which data became available first. It was found that there was no significant correlation between TSP and PFS and OS at any tumor evaluation time point. This is not very hard to understand as this approach only considers tumor regression at one single time point. The contribution of

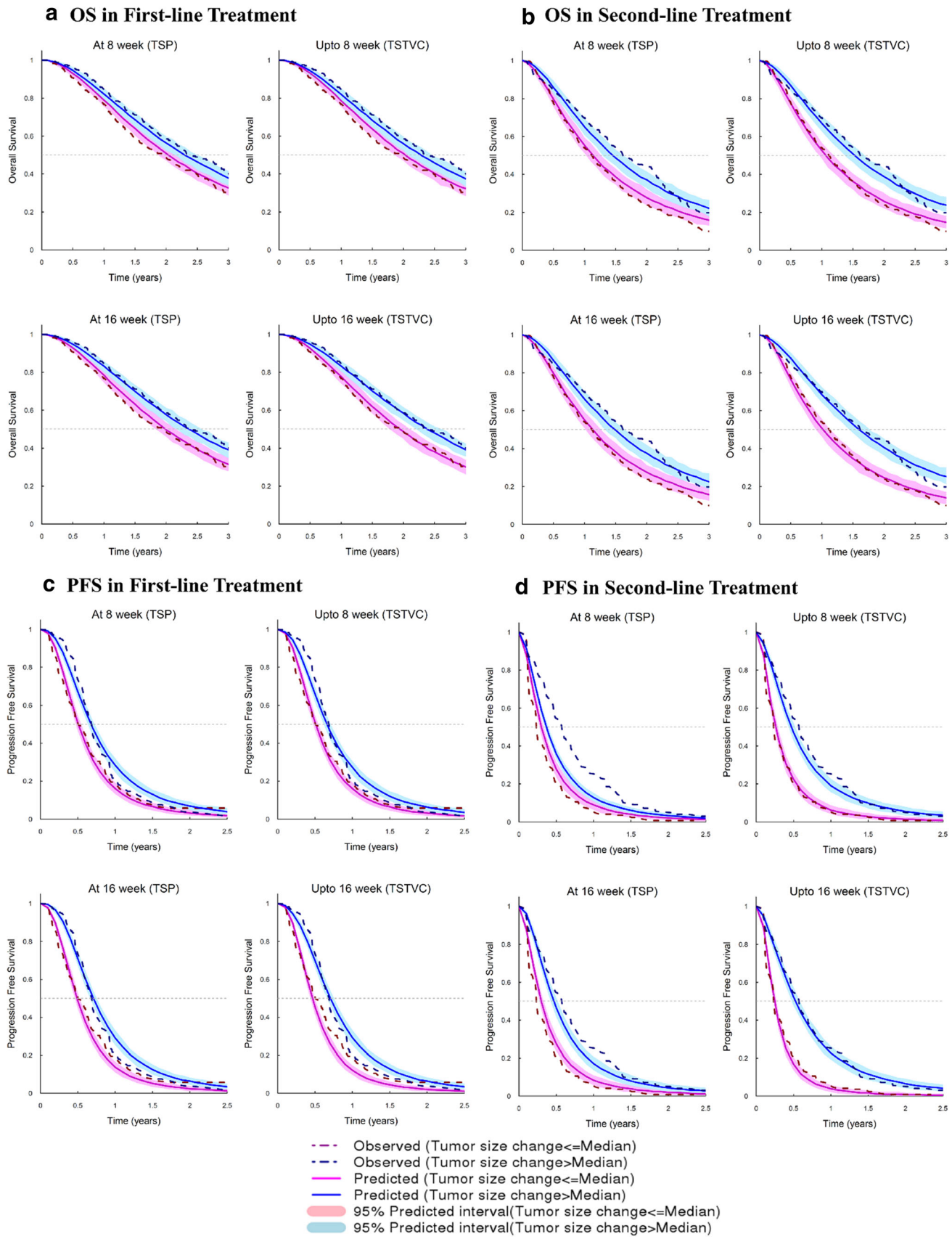


Fig. 3. Comparison of visual predictive check plots between models incorporating proportional tumor size change at 8 or 16 weeks (TSP) versus time-varying tumor size change up to 8 or 16 weeks (TSTVC) as covariates for OS in first-line (**a**) and second-line (**b**) treatment and for PFS in first-line (**c**) and second-line (**d**) treatment. The solid lines and shaded areas represent the median predicted survival curves and 95% confidence intervals, respectively. Broken lines represent the Kaplan-Meier survival curves based on observed data

Table III. Parameter Estimates of the Final Model for Overall Survival and Progression-Free Survival with First-Line and Second-Line Treatment

	First-line treatment		Second-line treatment	
Parameter[a]	Estimate	RSE%	Estimate	RSE%
OS				
λ	2.63	8.97	1.52	7.57
γ	2.10	2.09	1.96	3.49
β (ER = 0 [−] versus ER = 1 [+])	0.332	22.11	0.360	19.50
β (HER2 = 0 [−] versus HER2 = 1 [+])	–	–	0.400	20.20
β (ECOG = 0 versus ECOG > 1)	− 0.279	23.30	− 0.374	18.90
β (BSLD)	− 0.160	16.81	− 0.143	24.34
β (baseline AST)	− 0.213	15.31	− 0.188	17.50
β (TSP)	0.437	20.98	–[b]	–
β (TSTVC)	–[b]	–	0.793	16.14
PFS				
λ	0.394	7.72	0.382	6.13
γ	2.55	4.71	2.29	3.15
β (ER = 0 [−] versus ER = 1 [+])	0.292	18.63	–	–
β (Group = 0 ER[−] and HER2 [−] versus Group = 1 others)	–	–	− 0.450	13.58
β (ECOG = 0 versus ECOG > 1)	–	–	− 0.158	36.84
β (baseline AST)	− 0.0777	36.42	− 0.0892	26.68
β (TSP)	1.08	12.22	–[b]	–
β (TSTVC)	–[b]	–	1.88	8.83

Tumor size change (%) at 16 weeks was used as a covariate (TSP) in the final model for overall survival (OS) in first-line treatment and tumor size change (%) at 8 weeks for progression free survival (PFS). Tumor size change (%) up to 8 weeks was used as a time-varying covariate (TSTVC) with last prediction carried forward (LPCF) for time points beyond 8 weeks in the final models for OS and PFS in second-line treatment. In order to report the accurate β estimates for tumor size change in survival model, individual tumor size parameters obtained based on all available tumor size data were used to provide the tumor size prediction

AST aspartate aminotransferase (centered on 26 U/L), *BSLD* baseline sum of the longest diameter of the tumor (centered on 6.5 cm), *ECOG* Eastern Cooperative Oncology Group performance status (0 versus > 0), *ER* estrogen receptor (− versus +), *HER2* human epidermal growth factor receptor 2 (− versus +), *OS* overall survival, *PFS* progression-free survival

[a] λ and γ are scale and shape factors, respectively, in the log-logistic model. βs represent the coefficients relating a covariate to OS or PFS in the form of $S(t) = S_0\{t / \exp(\beta)\}$, where S is survival function and S_0 is baseline survival function

[b] β (TSTVC) was 0.528 and 1.12 for OS and PFS, respectively, in first-line treatment; β (TSP) was 0.402 and 0.848 for OS and PFS, respectively, in second-line treatment

tumor size change to predict survival endpoints could be same for patients who are on the regression course and progression course of the tumor dynamics when TSP is used without considering the tumor dynamics prior to that particular time point. The observation and this potential issue behind it led to the thought of using tumor size change as a time-varying covariate (TSTVC) in such evaluations. With the same Phase 2 dataset, TSTVC was shown as a significant covariate to improve the PFS and OS prediction. When sample size increased with additional data from other trials (first-line studies 2 and 3) becoming available, TSP contained enough information and became a significant covariate for PFS and OS in the survival model. In some situations, such as first-line treatment model, TSP approach performed as well as TSTVC approach. However, TSTVC approach provided significantly better predictions than the TSP model for second-line treatment. This could be due to the fact that the time course of tumor size reduction and then re-growth within shorter period of time could be more profound in second-line treatment due to more aggressive characteristics of disease compared to that in first-line patients. The tumor size-time course which was used by TSTVC approach carried more information and had stronger correlation with PFS/OS outcome as demonstrated by the higher β values of TSTVC approach compared to those of TSP approach, especially for second-line treatment (Table III).

For TSTVC approach, the tumor response beyond the observation time needs to be included in the model for survival predictions. For example, if tumor size changes up to 16 weeks are used for prediction, although the tumor size after 16 weeks is unknown, a value at each event time still needs to be included in the model for survival endpoint prediction. Ideally, the tumor model built based on data up to 16 weeks can also predict the tumor size change from 16 weeks to the event time for each patient well, allowing the model-predicted tumor size change to be used as the predictor. However, when this approach was compared to LPCF approach, the latter provided better prediction in all cases. Inspection of predicted versus observed tumor size data revealed that the prediction of tumor size time course post 16 weeks based on the tumor model built using only data up to 16 weeks can be way off for some patients which is likely the reason for the poor performance when predicted tumor size beyond evaluation time (16 weeks in this case) was used. In addition, other treatments were allowed beyond the PFS event or censoring while patients were followed up for OS in these trials, and tumor size data was normally unavailable during this time frame. The switch of treatments could

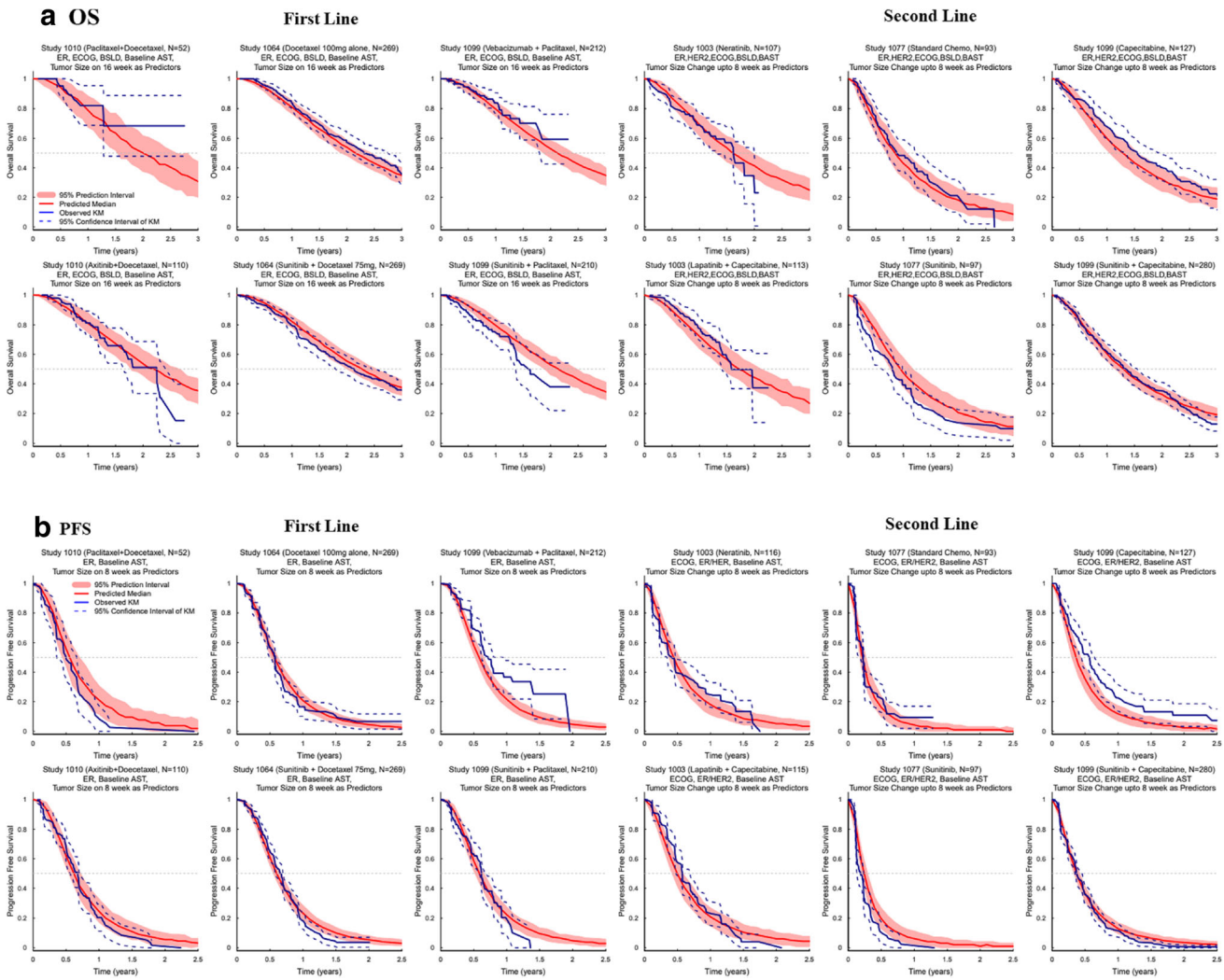


Fig. 4. Comparison of survival curves between simulated and observed data for each treatment arm. The red lines and pink-shaded areas represent the median survival curve and 95% confidence intervals (CIs), respectively, for simulated data. The solid and broken dark blue lines represent the Kaplan-Meier survival curve and its 95% CI based on observed data

certainly change the pattern of tumor size change from the previous study drug treatment, and this may contribute to the deviation of model prediction from the actual tumor size. Collectively, these may provide some explanation why using the last predicted value post observation time in the model is a more reliable approach.

Another question addressed in this analysis was what the optimal and earliest observation time of tumor response for a reliable prediction of survival endpoints is. The answer to this question varies with the tumor types and disease stages due to different average duration of survival as well as the frequency of tumor assessment in the clinical trials. Tate *et al.* reported that change in tumor size at week 8 was a useful predictor of OS in first-line metastatic breast cancer (8), and Claret *et al.* reported that tumor size change from baseline at week 7 was a significant predictor of OS in patients with metastatic colorectal cancer (2). In our analysis, a comprehensive research on the timing of tumor size observation was conducted by predicting the OS/PFS with tumor size change at and up to different time points in each test, and 8 weeks was selected as an early optimal time point for advanced breast cancer based on OFV changes and predictive performance shown in VPC plots for first-line PFS, second-line OS, and PFS while 16 weeks provided better prediction first-line PFS although the prediction using 8 weeks data was also reasonable.

Despite different types of breast cancer treatments included in the current analysis, there was only two targeted therapies (neratinib and lapatinib) while most of therapeutic agents were cytotoxic chemotherapies due to limited data available at the time of analysis. The analysis would be more comprehensive if more target therapies could be included especially in recent years small molecule targeted therapies have become one of the mainstay treatment options for breast cancer. Among patients with advanced breast cancer, approximately 17~37% of patients did not have measureable disease, *e.g.*, patients with only bone lesions (31). These patients had to be excluded from the current analysis evaluating tumor size change as a predictor for OS and PFS. In addition to significant increase in the SLD of target lesions, presence of new lesion is another sign of disease progression. Although the prediction for PFS was fairly reasonable based

on the final survival model especially in second-line treatment setting, there may still be room to improve the prediction for disease progression due to new lesions by identifying additional predictors, so the overall prediction for PFS could be further improved. Unfortunately, PR and HER2 could not be tested as covariates in first-line treatment due to relatively high percentage of missing values, which may compromise the model performance given the significant prognostic effect of HER2/PR expression in ABC.

CONCLUSION

In conclusion, tumor size change from baseline up to 8 weeks can be used to predict OS and PFS in first-line and second-line treatment of advanced breast cancer for the types of drugs studied, and tumor size change up to16 weeks can be used to achieve better predictions for OS in first-line treatment if feasible. While both the TSTVC model and the TSP model performed similarly for the prediction of OS and PFS in first-line treatment, the TSTVC model had better prediction for second-line treatment than the TSP model. The lack of treatment effect suggests that such models can be used to predict the efficacy outcomes for drug classes that have been studied in this analysis. These findings and the established models could contribute to drug development for advanced breast cancer treatment in various ways, including model-informed study design and early decision making in clinical drug development.

ACKNOWLEDGEMENTS

This study was sponsored by Pfizer. Editorial support was provided by Complete Healthcare Communications, LLC (West Chester, PA), a CHC Group company, and was funded by Pfizer.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of Interest K.P. and D.W. are employees of and shareholders of Pfizer Inc. W.S. was a former employee of Pfizer and a current shareholder of Pfizer Inc. H.-S.L. was a research fellow with Pfizer when this research was conducted.

REFERENCES

1. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, *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. 2009;86:167–74.
2. Claret L, Girard P, Hoff PM, van Cutsem E, Zuideveld KP, Jorga K, *et al*. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. J Clin Oncol. 2009;27:4103–8.
3. Lu JF, Claret L, Sutjandra L, Kuchimanchi M, Melara R, Bruno R, *et al*. Population pharmacokinetic/pharmacodynamic modeling for the time course of tumor shrinkage by motesanib in thyroid cancer patients. Cancer Chemother Pharmacol. 2010;66:1151–8.
4. Claret L, Lu JF, Sun YN, Bruno R. Development of a modeling framework to simulate efficacy endpoints for motesanib in patients with thyroid cancer. Cancer Chemother Pharmacol. 2010;66:1141–9.
5. Stein WD, Gulley JL, Schlom J, Madan RA, Dahut W, Figg WD, *et al*. Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. Clin Cancer Res. 2011;17:907–17.
6. Bruno R, *et al*. Simulations to assess phase II noninferiority trials of different doses of capecitabine in combination with docetaxel for metastatic breast cancer. CPT Pharmacometrics Syst Pharmacol. 2012;1(1):6.
7. Claret L, Bruno R, Lu JF, Sun YN, Hsu CP. Exploratory modeling and simulation to support development of motesanib in Asian patients with non-small cell lung cancer based on MONET1 study results. Clin Pharmacol Ther. 2014;95:446–51.
8. Tate SC, Andre V, Enas N, Ribba B, Gueorguieval I. Early change in tumor size predicts overall survival in patients with first-line metastatic breast cancer. Eur J Cancer. 2016;66:95–103.
9. US Food and Drug Administration. Innovation or stagnation: challenge and opportunity on the critical path to new medicinal products. Silver Spring, MD: US Department of Health and Human Services, US Food and Drug Administration; 2004.
10. European Medicines Agency. Road map to 2015. London, UK: European Medicines Agency; 2010.
11. Rugo HS, Stopeck AT, Joy AA, Chan S, Verma S, Lluch A, *et al*. Randomized, placebo-controlled, double-blind, phase II study of axitinib plus docetaxel versus docetaxel plus placebo in patients with metastatic breast cancer. J Clin Oncol. 2011;29:2459–65.
12. Bergh J, Bondarenko IM, Lichinitser MR, Liljegren A, Greil R, Voytko NL, *et al*. First-line treatment of advanced breast cancer with sunitinib in combination with docetaxel versus docetaxel alone: results of a prospective, randomized phase III study. J Clin Oncol. 2012;30:921–9.
13. Robert NJ, Saleh MN, Paul D, Generali D, Gressot L, Copur MS, *et al*. Sunitinib plus paclitaxel versus bevacizumab plus paclitaxel for first-line treatment of patients with advanced breast cancer: a phase III, randomized, open-label trial. Clin Breast Cancer. 2011;11:82–92.
14. Curigliano G, Pivot X, Cortés J, Elias A, Cesari R, Khosravan R, *et al*. Randomized phase II study of sunitinib versus standard of care for patients with previously treated advanced triple-negative breast cancer. Breast. 2013;22:650–6.
15. Crown JP, Diéras V, Staroslawska E, Yardley DA, Bachelot T, Davidson N, *et al*. Phase III trial of sunitinib in combination with capecitabine versus capecitabine monotherapy for the treatment of patients with pretreated metastatic breast cancer. J Clin Oncol. 2013;31:2870–8.
16. Martin M, Bonneterre J, Geyer CE Jr, Ito Y, Ro J, Lang I, *et al*. A phase two randomised trial of neratinib monotherapy versus lapatinib plus capecitabine combination therapy in patients with HER2+ advanced breast cancer. Eur J Cancer. 2013;49:3763–72.
17. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. Br J Cancer. 2003;89:431–6.
18. Patel K, Kay R, Rowell L. Comparing proportional hazards and accelerated failure time models: an application in influenza. Pharm Stat. 2006;5:213–24.
19. Keiding N, Andersen PK, Klein JP. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. Stat Med. 1997;16:215–24.
20. Hougaard P. Fundamentals of survival data. Biometrics. 1999;55:13–22.
21. Hougaard P, Myglegaard P, Borch-Johnsen K. Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. Biometrics. 1994;50:1178–88.
22. Laird AK. Dynamics of tumor growth. Br J Cancer. 1964;13:490–502.
23. Efficace F, Biganzoli L, Piccart M, Coens C, van Steen K, Cufer T, *et al*. Baseline health-related quality-of-life data as prognostic

factors in a phase III multicentre study of women with metastatic breast cancer. Eur J Cancer. 2004;40:1021–30.
24. Fisher B, Redmond C, Fisher ER, Caplan R. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. 6, 1076–1087 (1988).
25. Samaan NA, Buzdar AU, Aldinger KA, Schultz PN, Yang KP, Romsdahl MM, *et al*. Estrogen receptor: a prognostic factor in breast cancer. Cancer. 1981;47:554–60.
26. Swenerton KD, Legha SS, Smith T, Hortobagyi GN, Gehan EA, Yap HY, *et al*. Prognostic factors in metastatic breast cancer treated with combination chemotherapy. Cancer Res. 1979;39:1552–62.
27. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989;63:181–7.
28. Bruno R, Claret L. On the use of change in tumor size to predict survival in clinical oncology studies: toward a new paradigm to design and evaluate phase II studies. Clin Pharmacol Ther. 2009;86:136–8.
29. Bruno R, Lu JF, Sun YN, Claret L. A modeling and simulation framework to support early clinical drug development decisions in oncology. J Clin Pharmacol. 2011;51:6–8.
30. Wang E, Kang D, Bae KS, Marshall MA, Pavlov D, Parivar K. Population pharmacokinetic and pharmacodynamic analysis of tremelimumab in patients with metastatic melanoma. J Clin Pharmacol. 2014;54:1108–16.
31. Ahn SG, Lee HM, Cho SH, Lee SA, Hwang SH, Jeong J, *et al*. Prognostic factors for patients with bone-only metastasis in breast cancer. Yonsei Med J. 2013;54:1168–77.