Citation: CPT Pharmacometrics Syst. Pharmacol. (2016) 5, 274–282; doi:10.1002/psp4.12078

ORIGINAL ARTICLE

Clinical Trial Simulations From a Model-Based Meta-Analysis of Studies in Patients With Advanced Hepatocellular Carcinoma Receiving Antiangiogenic Therapy

ML Zierhut[1]*†, Y Chen[2], YK Pithavala[2], DJ Nickens[1], O Valota[3] and MA Amantea[1]

A mixed effect model describing median overall survival (mOS) in patients with advanced hepatocellular carcinoma (aHCC) treated with antiangiogenic therapy (AAT) was developed from literature data. Data were extracted from 59 studies, representing 4,813 patients. The final model included estimates of mOS after AAT (8.5 months) or placebo (7.1 months) administration. The mOS increased 21% when the AAT was sorafenib (SOR) or 42% when locoregional therapy was coadministered. The mOS decreased when patients received prior systemic therapy (↓7%) or concomitant chemotherapy (↓4%) or the percentage of patients with hepatitis B increased (↓∼0.4%/%). Clinical trial simulations of a phase II comparative trial predicted an mOS ratio (placebo:AAT) of 0.687 or 0.831, with a 65% or 22% probability of demonstrating superiority, for SOR or other AATs, respectively. Additionally, the 95% confidence interval (CI) of the simulated median mOS ratio for non-SOR AATs was similar to the 95% CI of the hazard ratio (HR) observed in the trial.
CPT Pharmacometrics Syst. Pharmacol. (2016) **5**, 274–282; doi:10.1002/psp4.12078; published online 15 May 2016.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC? ☑ Systematic reviews have been published for sorafenib and other AATs in HCC, but no rigorous MBMA of mOS in a population receiving AAT was found in the literature. • WHAT QUESTION DID THIS STUDY ADDRESS? ☑ This analysis was conducted to develop a model describing the range of mOS values reported in aHCC studies with systemic AATs. Clinical trial simulations were performed to help interpret the results of a phase II trial and to guide future study designs. • WHAT THIS STUDY ADDS TO OUR KNOWLEDGE ☑ This analysis utilized 59 clinical studies, representing 4,813 patients with aHCC, to identify seven predictors of mOS and to quantify within and between trial variability. This analysis also highlighted the ability to perform MBMA-based clinical trial simulations. • HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS ☑ The approach used here could be adapted to improve the efficiency of any drug development program. It adds to the growing body of work demonstrating the utility of MBMA in real time clinical development of investigational agents.

Hepatocellular carcinoma (HCC) is a highly vascular tumor in which vascular recruitment and invasion greatly contribute to pathogenesis. The vascular endothelial growth factor (VEGF) is thought to have an important role in HCC angiogenesis; its expression has been confirmed in this disease and has been associated with a poor prognosis.[1] Agents that inhibit angiogenesis pathways may increase the therapeutic options for patients with HCC with altered liver function, and may offer a potentially better safety profile in comparison with chemotherapy agents. Systemic antiangiogenic agents, including sorafenib (SOR) and bevacizumab, have shown antitumor activity in HCC.[2,3] Sorafenib is currently the only antiangiogenic therapy (AAT) approved to treat advanced HCC (aHCC).

Axitinib (AG 013736; Inlyta) is an oral, potent, and selective inhibitor of VEGF receptors 1, 2, and 3. Axitinib has been approved as second-line therapy for advanced renal cell carcinoma in more than 70 countries (actual indication varies). Based on the activity of several other VEGF inhibitory agents in phase II HCC studies, and the nonclinical activity of axitinib in HCC animal models,[4] there was a rationale for testing the safety and efficacy of axitinib in patients with aHCC. Consequently, a phase II clinical trial was conducted to compare the efficacy of axitinib plus best supportive care (BSC) to placebo plus BSC in patients with aHCC who had failed one prior AAT (NCT01210495).[5] The primary end point was overall survival (OS).

The model-based meta-analysis (MBMA) described here was based on previously published information about trials in patients with aHCC treated with systemic antiangiogenic agents (not axitinib) that reported median overall survival (mOS). Systematic reviews of antiangiogenic agents in aHCC trials have been published;[6,7] however, the current MBMA was intended to be a more inclusive and thorough analysis using clinical data available through late 2012, with the additional incorporation of fixed and random effects.

[1]Pharmacometrics, Clinical Pharmacology, Pfizer, La Jolla, California, USA; [2]Oncology, Clinical Pharmacology, Pfizer, La Jolla, California, USA; [3]Oncology, Clinical Development, Pfizer, Milan, Italy. *Correspondence: ML Zierhut (matt.zierhut@gmail.com)
Received 16 November 2015; accepted 21 March 2016; published online on 15 May 2016. doi:10.1002/psp4.12078
†Clinical Pharmacology and Pharmacometrics, The Janssen Pharmaceutical Companies of Johnson & Johnson, San Diego, California, USA.

Table 1 Summary of indicator variables

Treatment characteristic	No. of arms	Study characteristic	No. of trials
PBO	6	Blinded	10
SOR	26	Phase II or III	43
Other AAT agent[a]	36	Single arm	49
Concomitant LOC	7	Randomized	9
Concomitant CTx	13	Asian study site	22
Combination therapy[b]	24		
All second line patients	7		
All first line patients	43		
All	68	All	59

AAT, antiangiogenic therapy; CTx, chemotherapy; LOC, locoregional; PBO, placebo; SOR, sorafenib.
[a]Fourteen other AATs (ranging from 1–9 arms per AAT).
[b]One triple therapy arm.

This MBMA was performed to help gain insight on the overall benefit of AAT in aHCC and to inform development decisions for axitinib as a therapeutic option for aHCC. Specifically, the primary objectives of this analysis were: (1) to quantify the range of mOS values observed in studies with systemic AAT treatment in patients with aHCC; (2) to identify significant predictors as sources of variability of mOS; and (3) to better understand the probability of demonstrating axitinib superiority in a phase II trial (NCT01210495).

RESULTS

Data summary

The initial literature search identified 350 publications. **Supplementary Figure S1** summarizes how sources were subsequently filtered during the literature review and data extraction process. The resulting final dataset contained data from 59 studies with 68 total treatment arms, representing a total of 4,813 patients.

Table 1 summarizes the indicator variables in the analysis dataset by treatment arm and trial, whereas **Table 2** summarizes the continuous variables in the analysis dataset.

Final model

The initial fixed effect model structure for this analysis incorporated only a single parameter (the intercept). An additional parameter was added to distinguish between arms receiving AAT and placebo (PBO), as AAT and PBO are mutually exclusive and collectively exhaustive indicator variables. All patients were assumed to have received BSC in addition to any listed treatment, thus, BSC is implicitly built into the model.

Forward covariate selection resulted in three variables selected in the following order: SOR treatment indicator, concomitant locoregional (LOC) therapy indicator, and percent of population with the hepatitis B virus (HBV). LOC therapy consisted of transarterial chemoembolization, SIR-spheres, or cryoablation. Prior systemic therapy (PTx) and concomitant chemotherapy (CTx) were considered to be clinically relevant variables in this target population, and, thus, indicators for these characteristics were included in the model as additive covariates, even though they did not meet the significance criteria. No covariates were removed from the model during the backward elimination step.

The final model had the form:

$$\ln\left(\mathrm{mOS}_{ij}\right) = \left(\theta_{AAT}\cdot \mathrm{AAT}+\theta_{PBO}\cdot \mathrm{PBO}+\eta_j\right)+\theta_{SOR}\cdot \mathrm{SOR} +\theta_{LOC}\cdot \mathrm{LOC}+\theta_{HBV}\cdot \mathrm{HBV}+\theta_{PTx}\cdot \mathrm{PTx}+\theta_{CTx}\cdot \mathrm{CTx}+\varepsilon_{ij}\cdot \mathrm{SE}_{ij};$$

where $\ln(\mathrm{mOS}_{ij})$ is the response in the i^{th} treatment arm of the j^{th} trial; η is a random effect on study, having a normal distribution with mean 0 and variance ω^2; SE is the reported standard error of the response; and ε is the residual error component, having a normal distribution with mean 0 and variance σ^2. The subscript on the fixed effect parameter (θ) indicates the covariate associated with that estimated effect; for example, θ_{SOR} represents the shift in ln(mOS) associated with SOR treatment.

Table 2 Summary of continuous variables

Variables	Value
Mean age (y)	60.2 (6.8)
Reported in 83% of arms	60 [47, 75]
Percent male	78.6 (9.7)
Reported in 79% of arms	80.1 [55, 94.7]
Publication year	2010.6 (1.70)
Reported in 100% of arms	2011 [2005, 2012]
Percent with Child-Pugh A disease	82.7 (16.1)
Reported in 80% of arms	83.9 [7.4, 100]
Percent with Child-Pugh B disease	13.7 (15.5)
Reported in 79% of arms	8.5 [0, 92.6]
Percent with aHCC	98.1 (6.0)
Reported in 99% of arms	100 [69, 100]
Percent with HBV	38.2 (26.1)
Reported in 70% of arms	34.0 [4.8, 100]
Percent with prior vascular invasion	35.2 (16.7)
Reported in 43% of arms	31.9 [0, 100]
Percent with ECOG = 0	43.1 (17.4)
Reported in 58% of arms	36.3 [0, 100]
Percent with ECOG <2	90.8 (9.5)
Reported in 73% of arms	94.4 [76.8, 100]
Percent with prior CTx	20.5 (31.6)
Reported in 63% of arms	14.1 [0, 100]
Percent with prior systemic tx	16.6 (32.0)
Reported in 90% of arms	0 [0, 100]
No. of patients	70.8 (100.1)
Reported in 100% of arms	42 [10, 544]
mOS (months)	6.6 (3.3)
Reported in 100% of arms	9.4 [4.2, 20.8]
ln(mOS)[a]	2.18 (0.36)
Reported in 100% of arms	2.24 [1.44, 3.04]
SE of ln(mOS)[b]	0.186 (0.11)
Reported in 82% of arms	0.183 [0.045, 0.61]

Mean (SD).
Median [min, max].
aHCC, advanced hepatocellular carcinoma; CTx, chemotherapy; ECOG, Eastern Cooperative Oncology Group; HBV, hepatitis B virus; mOS, median overall survival; tx, treatment.
[a]Response variable used in model (dependent variable).
[b]Used for weighting of residual error term, ε.

Table 3 Summary of final model parameter estimates

Parameter	Estimate	95% CI[a]
$\exp(\theta_{AAT})$ [mo]	8.49	(7.72–9.36)
$\exp(\theta_{PBO})$ [mo]	7.06	(6.28–7.86)
$\exp(\theta_{SOR})$ [%]	121	(114–128)
$\exp(\theta_{LOC})$ [%]	142	(122–165)
θ_{HBV} [d.u./%][b]	−0.00418	$(-7.02, -1.34) \times 10^{-3}$
$\exp(\theta_{PTx})$[c] [%]	93.3	(74.0–118)
$\exp(\theta_{CTx})$[c] [%]	96.0	(80.7–114)
ω[d] [shrinkage, %]	0.216 [21.2%]	(0.166–0.266)
σ[d] [shrinkage, %]	1 (fixed) [41.5%]	–

AAT, antiangiogenic therapy; CI, confidence interval; CTx, chemotherapy; HBV, median-centered percent of hepatitis B-positive patients; LOC, locoregional therapy; PBO, placebo; PTx, prior systemic therapy; SOR, sorafenib.
[a]95% CI calculated as parameter estimate ± 1.96·SE.
[b]Parameter in log-domain, thus unitless (d.u. = dimensionless unit).
[c]Parameter forced into model.
[d]ω and σ are reported as SDs of random variables η and ε, respectively.

Table 3 shows the final MBMA parameter estimates. These parameters show that, for a typical population (no prior therapy, 34% of patients with HBV), monotherapy with a non-SOR AAT leads to an average 1.43-month benefit in mOS compared to PBO (8.49 vs. 7.06 months). This benefit is increased by 21% if the AAT is SOR and by 42% if the population receives concomitant LOC therapy. The mOS decreases by ~0.4% for every 1% increase in HBV-positive patients. Additionally, mOS decreases if the population has received prior therapy (6.7% decrease) or if the population receives concomitant CTx (4% decrease); however, these parameters were not significant, and the 95% confidence intervals (CIs) of their estimates spanned both beneficial and detrimental effects. Additionally, with σ fixed to one, no other fixed effect parameter estimate changed by >10% (range, −9.7 to 5.8%) and ω decreased by ~20%, compared to estimating σ at 0.568.

Model evaluation

Selected residual-based diagnostics are shown in **Figure 1**. The individual predictions and observed values seem to fall close to the line of unity, the magnitudes of the population weighted residual (WRES) values are all less than three, with no apparent structure relative to population predicted values and the quantile-quantile plots suggest that the residual variability and between-study variability terms are well approximated using a normal distribution.

Simulation-based diagnostics are shown in **Figure 2**. **Figure 2a** summarizes the observed OS for each trial arm (median and 95% CI) compared to the distribution of respective mOS derived from the simulations (median and 95% prediction interval). **Figure 2b** shows the results from deterministic meta-analyses using the observed dataset compared to the distribution of results from the simulated datasets. Similar to the residual-based diagnostics, these plots suggest that the model is adequately describing the observed data.

All pairwise correlations between parameter estimates were <0.8 and the condition number (ratio of largest to smallest eigenvalues) was <20, suggesting model structure was adequate to describe the observed data. Publication bias was assessed via funnel plots of model residual values (not shown), as well as with the metabias() function within the "meta" library of the R programming language. These were assessed for all data (P value = 0.22), as well as separately for SOR-containing treatment arms (P value = 0.22) and single arm trials (P value = 0.70). No significant publication bias was found.

Clinical trial simulations

Because it was not known if SOR or other AATs would be more appropriate to predict how the target patient population would respond to axitinib, two sets of simulations were performed. One simulation applied the effects of a non-SOR AAT, whereas the other incorporated effects of SOR. All other model parameters were identical between the two sets of simulations. **Figure 3** shows the resulting distributions for non-SOR therapy (**Figure 3a,c,e**) and SOR therapy (**Figure 3b,d,f**).

Based on these simulations, in a population similar to this phase II trial, mOS (95% CI) is expected to be 6.16 (3.58–10.6), 7.40 (4.48–12.3), and 8.95 (5.38–14.9) months after treatment with BSC plus PBO, non-SOR AAT, and SOR, respectively. Additionally, the ratio of mOS values may be used to approximate the expected OS hazard ratio (HR), assuming the survival curve for each treatment arm follows an exponential distribution. This assumption results in HRs (95% CI) of 0.831 (0.606–1.14) and 0.687 (0.502–0.943) for PBO compared to non-SOR AAT and SOR, respectively. Furthermore, these simulations suggest that if axitinib were found to be similar to other non-SOR AATs, this phase II trial would have approximately a 22% probability of demonstrating axitinib's superiority to PBO. If axitinib were similar to SOR, the trial would have approximately a 65% chance of demonstrating superiority. The "true" mOS ratio (95% CI) of PBO compared to AAT was simulated to be 0.831 (0.774–0.891) and 0.687 (0.640–0.738) for non-SOR therapy and SOR therapy, respectively, suggesting any AAT (SOR or not) is superior to PBO. In order for this trial to have an 80% probability of significantly demonstrating superiority, simulations suggest that at least 1,221 patients (814 active, 407 PBO) or 282 patients (188 active, 94 PBO) would be needed if axitinib were similar to a non-SOR AAT or SOR, respectively.

Actual OS results from this phase II trial were available after this MBMA was performed and were presented at the European Society for Medical Oncology annual meeting in 2014.[8] In this trial, patients in the axitinib arm had an mOS (95% CI) of 12.7 (10.2–14.9) months, whereas patients receiving PBO had an mOS (95% CI) of 9.7 (5.9–11.8) months. This difference was not statistically significant, with a HR of 0.870 (95% CI = 0.620–1.22; P = 0.211). Although the observed mOS values do not seem to be similar to the simulations, the mOS ratio of the non-SOR simulation is similar to the HR observed in the trial. Specifically, the median 2.5th and 97.5th percentiles of the simulated trial mOS ratio (0.609 and 1.13, respectively) were similar to the observed 95% CI of the HR (0.620–1.22) and the observed HR (0.870) is contained within the 95% CI of the "true" mOS ratio (0.774–0.891).

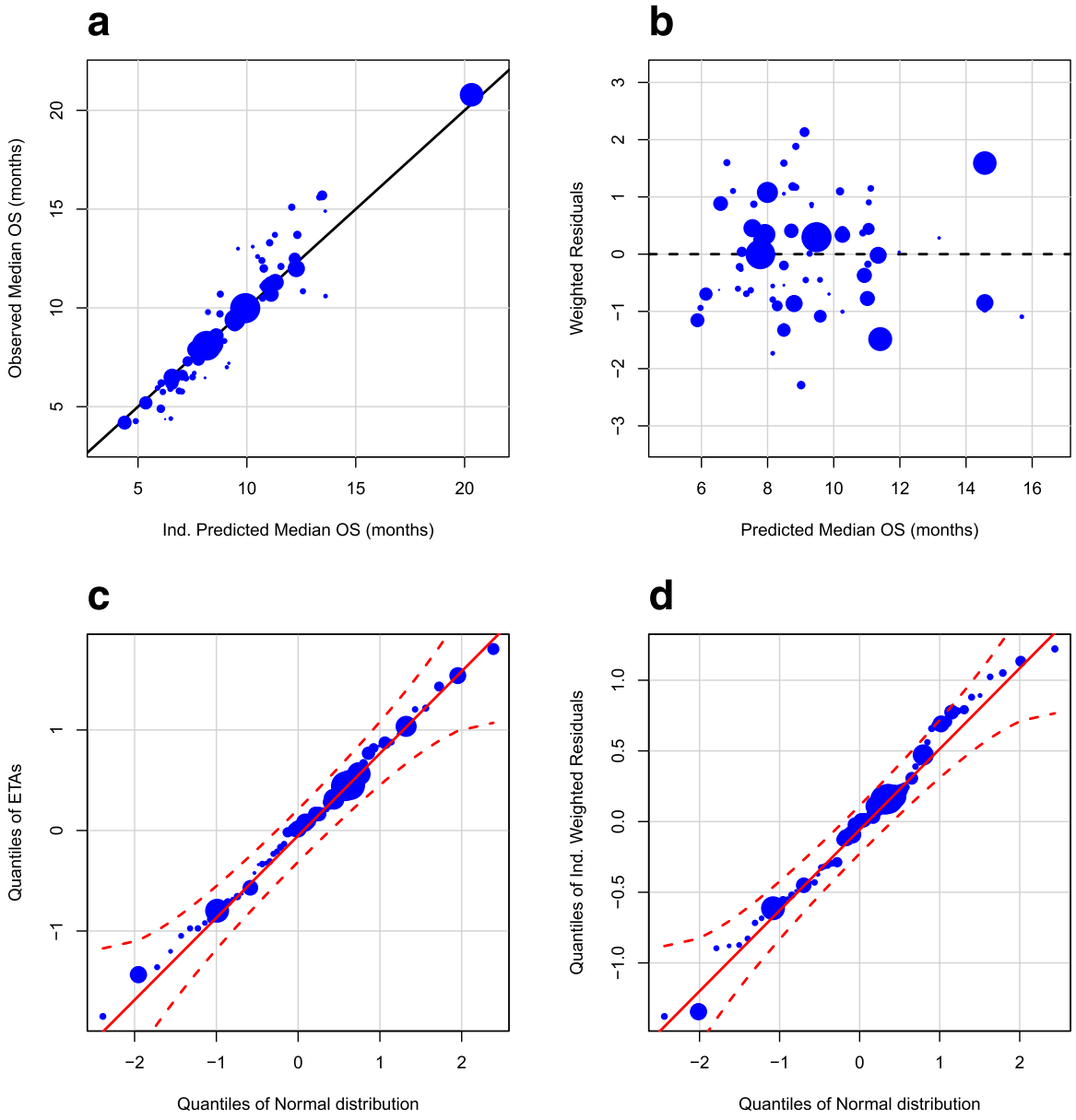

Figure 1 Residual-based diagnostic plots for the final model showing (**a**) observed response vs. predicted individual (treatment arm) response, (**b**) population weighted residual (WRES) vs. population predicted response and quantile-quantile plots of (**c**) between-subject variability random effects (η), and (**d**) population WRES. Symbol size represents the weighting by reported SE. Red dashed lines in (**c**) and (**d**) represent the 95% confidence interval for the pointwise confidence envelope.

DISCUSSION

The purpose of this analysis was to gain a deeper understanding of the benefit of systemic AAT in patients with aHCC, based on aggregate data in published literature. A mixed effects model was developed to describe the variability of mOS reported in various literature sources. This model was found to adequately describe the literature data and was subsequently used to simulate mOS in a phase II trial in order to better understand the probability of demonstrating axitinib superiority to PBO.

The final model consisted of seven fixed effect parameters (θ_{AAT}, θ_{PBO}, θ_{SOR}, θ_{LOC}, θ_{HBV}, θ_{PTx}, and θ_{CTx}) and two random effect parameters (η and ε). Both residual- and simulation-based diagnostics indicated that this model adequately described the observed data. Five of the fixed effect parameters were indicator variables for current treatment characteristics (PBO, AAT, SOR, LOC, and CTx), and two parameters described the differences in mOS based on baseline population characteristics (HBV and PTx). Additionally, five of the seven fixed effect parameters were shown to be statistically significant based on the threshold defined in the backward covariate elimination step. The remaining two parameters (PTx and CTx) were forced into the model based on historical clinical relevance in the target patient population. These forced-in parameters were not

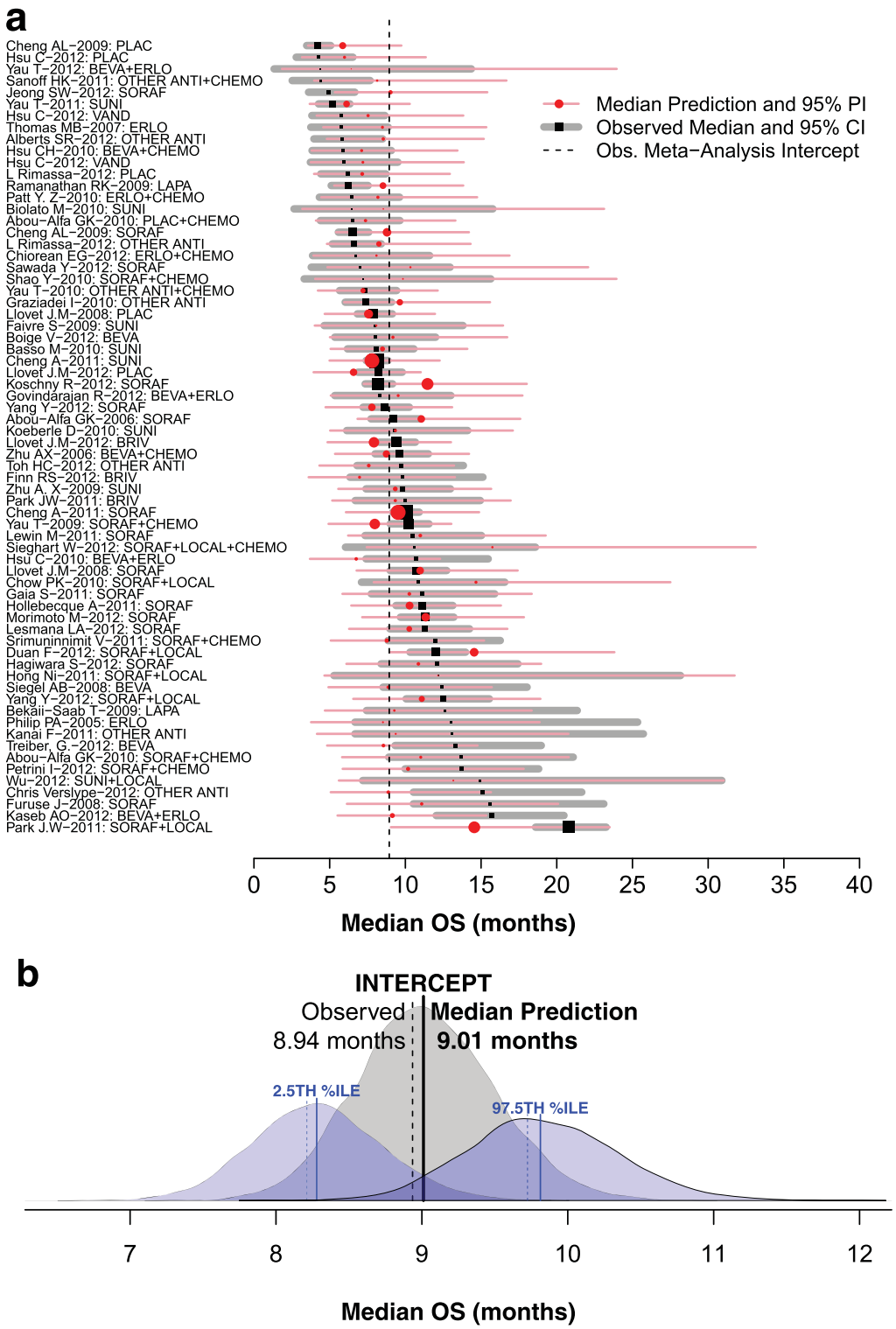

Figure 2 Simulation-based diagnostic plots for the final model showing (**a**) distributions of simulated data compared to observed data and (**b**) distributions of deterministic meta-analyses from simulated data and observed data. For (**a**), each datapoint is labeled by the source author, publication year, and treatment administered; symbol size represents the weighting by reported SE; red symbols represent simulated data, black symbols represent observed data. For (**b**), dashed lines represent deterministic meta-analysis results based on observed data, whereas solid lines and distributions represent results based on simulated data; black color represents the median result of the deterministic meta-analysis, with simulated distribution in gray; blue color represents the resulting 2.5th and 97.5th percentiles, with simulated distributions in light blue. CI, confidence interval; OS, overall survival; PI, prediction intervals.

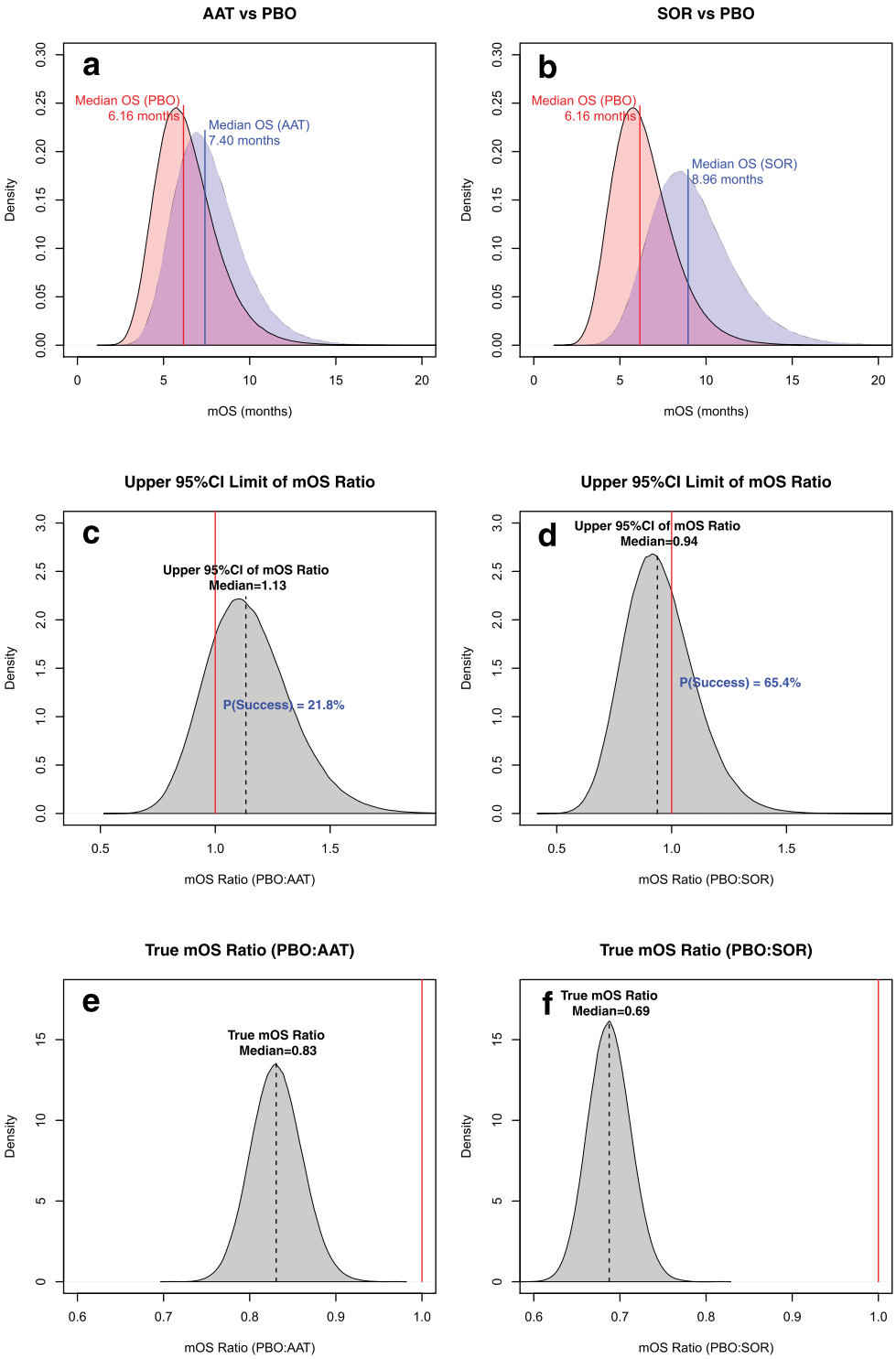

Figure 3 Results from phase II clinical trial simulations. Distributions are shown for both non-sorafenib (SOR) therapy (antiangiogenic therapy (AAT); (**a**, **c**, and **e**) and SOR therapy (**b**, **d**, and **f**), summarized as the distributions of median overall survival (mOS) in each treatment arm within the trial (**a** and **b**), the upper bound on the 95% confidence interval (CI) of the ratio of mOS values between treatment arms (placebo [PBO]:active) (**c** and **d**) and the expected ratio of mOS values for a trial with infinite patients ("true" mOS ratio; **e** and **f**). For (**a**) and (**b**), red represents placebo arm and blue represents active arm. For (**c**) and (**d**), the dashed vertical line represents the median upper limit of the 95% CI, the red vertical line represents the significance threshold and the probability of success is calculated as the relative area of the distribution to the left of the significance threshold. For (**e**) and (**f**), the dashed vertical line represents the "true" mOS ratio between the two arms, where the influence of number of patients in the trial has been removed.

highly influential, but they were retained to account for any potential difference in the clinical trial simulations. Because these two parameters were not significant, one should be cautious when drawing conclusions from their apparent effects, particularly when interpreting the apparent negative impact of concomitant CTx. It is also important to remember, however, that the apparent negative impact of CTx is in addition to the benefit of AAT. That is, it may be possible that CTx removes some benefit of AAT when administered concomitantly.

The five parameters that describe a patient's current treatment (PBO, AAT, SOR, LOC, and CTx) may suggest what an optimal treatment plan may look like. It seems that AAT added to BSC is more beneficial than BSC alone (or BSC + PBO). Specifically, SOR seems to have the most benefit of the antiangiogenic agents explored here. Additionally, concomitant locoregional therapy seems to be highly beneficial, whereas CTx may not have any benefit when administered with AAT. The two parameters that describe the baseline status of patients in a trial (HBV and PTx) may be used to aid in prognoses. Trials that have more patients with HBV tend to have a worse mOS, suggesting that a patient with HBV may be less likely to respond to AAT compared to a patient without HBV. Finally, whether patients in a trial had received prior systemic therapy or not does not seem to be a predictive factor for mOS.

Clinical trial simulations suggested that, for a patient population with aHCC, axitinib may have a similar efficacy (mOS benefit) to non-SOR AATs, with a low probability of demonstrating superiority to PBO in this phase II trial. Simulations also suggested that the trial would need 1,221 patients to have an 80% probability of demonstrating superiority. If axitinib were similar to SOR, 282 patients would be sufficient to achieve 80% probability of demonstrating superiority, highlighting the 21% increase in mOS compared to non-SOR AATs. However, after removing the influence of patient number, the "true" probability that a non-SOR AAT was superior to PBO approaches 100%.

There are some important limitations to the MBMA methodology as applied to clinical trial simulations. First, the dataset was comprised of summaries of observed data, aggregated to the level of the trial arm; thus, each data record represented an entire trial arm. Covariate relationships at this aggregate level may not be the same as at a patient level, and important relationships may have not been included in the final model. Second, there were not many treatment arms in which the population received PBO ($n = 6$) or the entire population had received prior therapy ($n = 7$). These were important factors for the clinical trial simulations, and their true impact may not have been fully accounted for. Third, most of the studies included in this dataset were single arm trials (49/59); thus, simulations of comparative trials based on this model may have limited inferential value. Despite these limitations, MBMA does demonstrate the ability to gain insight into a development program prior to obtaining any direct clinical results. Results from MBMA-based clinical trial simulations may be beneficial to guide trial design, but may be most informative in interpreting clinical trial results.

The observed results from the phase II trial of interest did not seem to be representative of prior clinical trials of AATs in similar populations when comparing mOS values in the trial arms. One possible explanation for this perceived discrepancy was that BSC has been improving over time, leading to an improvement in mOS of placebo arms as well as active treatment arms (with a constant relative benefit). This may help explain why the clinical trial simulation results showed an mOS ratio (and 95% CI) similar to the HR observed in this trial, but lower mOS values. Publication year was prospectively investigated as a potential predictor of mOS and was not found to be significant (change in minimum objective function [ΔMOF] = -0.449). This does not necessarily mean that BSC has not been improving over time, just that the dataset used to develop this model does not show evidence of this relationship. Nevertheless, this model can be continuously updated as new data become available to improve predictive capabilities.

Observed results from this phase II trial suggested that patients at Asian sites may benefit more from axitinib than patients at non-Asian sites.[8] Asian geography was prospectively investigated as a potential covariate in this MBMA, but was not found to be significant. After the trial showed a positive trend in OS due to Asian geography, most notably in the PBO arm, trial location was investigated further as a covariate unique to PBO treatment. The trial reported mOS for active vs. placebo arms as 13.5 vs. 6.3 months (HR = 0.809) and 12.3 vs. 11.2 months (HR = 0.971) for patients in Asian and non-Asian sites, respectively. Although Asian geography was not found to be a significant covariate on PBO treatment in the MBMA, a positive trend did exist (ΔMOF = -2.63). This relationship indicated a similar trend to that observed in the trial, with the model-predicted mOS ratio for Asian and non-Asian geographies at 0.821 and 0.902, respectively.

The MBMA performed here was able to adequately describe observed mOS values in 59 published clinical studies. Simulations based on this model were used to estimate the probability of a phase II trial's success and to help put observed trial results in context of previously published trials. MBMA-based simulations could further enhance drug development in multiple ways. Specifically, this approach could be utilized to more efficiently design prospective clinical trials when clinical data are not yet available or prevalent. This approach could also be used to create a virtual comparator arm, or a hybrid comparator arm, incorporating the MBMA as a Bayesian prior, and thus reducing the number of patients needed to achieve the necessary statistical power. Finally, outputs from this approach could be integrated into other, non-clinically focused models, such as cost-effectiveness models or financial forecasting models, potentially leading to improved commercial and financial strategies.

METHODS

Literature search and review

The following databases were included during the literature search: OVID Medline, Embase, Embase Alerts, and Medline in Process. Studies with patients with aHCC were included regardless of randomization, blinding, or line of treatment, provided that AAT was administered (both single agent and combination regimens included). PBO arms were included only if the active arm within that trial received AAT. Treatments of interest were inhibitors of VEGF receptors or

ligands and inhibitors of epidermal growth factor receptor. The primary end point of interest was mOS.

Retrospective analyses, study protocols, reviews, case studies, trials without AAT, duplicate trials, inaccessible publications, and study arms not reporting mOS were removed from the initial list of sources. Treatment arms were also excluded from the analysis dataset if they had population characteristics that were deemed to be potentially confounding to a model used to simulate the phase II trial of interest.[5] Additional relevant publications not found in the original search were subsequently added. The resulting list of publications was sent to an external vendor (GVK Biosciences Private Limited, Hyderabad, India) for data extraction of prospectively identified variables.

Data processing

After data extraction, a quality control review was performed to ensure the accuracy of data (e.g., numeric end points used in the analysis were compared for accuracy against the literature sources).

Data manipulations and creation of derived variables were performed using the R programming language, with "gdata" and "reshape" libraries (version 2.15.2; R Foundation for Statistical Computing, Vienna, Austria). Variables were removed from the analysis dataset if values were not reported for at least 65% of the patients represented in the dataset. A 65% threshold was empirically chosen to include the most data with a minimal potential impact to model development. Variables were imputed, when missing, as the average among the treatment arms reporting values, weighted by patient number.

The mOS was assumed to have a log-normal distribution; thus, the modeled response variable was ln(mOS). Residual variability in the model was weighted by the SE of ln(mOS) in each treatment arm; however, SE was not directly reported in the literature. Data sources reported variability in OS as 95% CI or ranges, or they did not report any OS variability. The SE was derived for each (i^{th}) treatment arm according to the following computations:

- If 95% CI was reported : $SE_{\ln(mOS),i} = \frac{p_{97.5,i} - p_{2.5,i}}{3.92}$

  where p_x represents the natural log (ln) of the reported x^{th} confidence limit and 3.92 represents the number of SEs spanned by the 95% CI.

- If range (min, max) was reported : $SE_{\ln(mOS),i} = \frac{\ln(\max_i) - \ln(\min_i)}{X \cdot \sqrt{N_{patients,i}}}$,

  where $X = 3$, 4, 5, or 6 if $N_{patients}$ was <10, between 11 and 25, between 26 and 100, or >100, respectively.[9]

- If no variability was reported : $SE_{\ln(mOS),j} = \sqrt{\frac{median^{\Delta}(SD^2_{\ln(mOS)})}{N_{patients,j}}}$,

  where $SD_{\ln(mOS)} = SE_{\ln(mOS)} \cdot \sqrt{N_{patients}}$, $median^{\Delta}$ represents the median of only the treatment arms with calculable SE and j indexes all treatment arms where no OS variability was reported.

Model building

A model with fixed and random effects was used to describe the range of mOS values. This analysis used published aggregate data (at the study arm level) to identify significant predictors of mOS. These predictors, or covariates, were added as fixed effects (θ_n) to reduce the random variability in the range of responses. Model building was performed using the R programming language with "nlme" library and final model parameters were estimated using NONMEM software, first order conditional estimation method with interaction (version 7.2; Icon Development Solutions, Ellicott City, MD).

Initial model structure utilized a single fixed effect parameter representing the typical value of all responses (the intercept; θ_1). Additive residual variability was weighted by the SE reported in each study arm. An additional additive random effect was incorporated into the base model to account for between-study variability on the intercept.

Covariate effects were necessary to describe any difference in responses between treatment arms in the same study. Because the response variable was log-transformed, the additive covariates were interpreted as proportional to mOS. All continuous covariates were centered to the median value to simplify interpretation of parameter estimates. The following equation demonstrates how the fixed and random effects were incorporated:

$$\ln(\mathrm{mOS}_{ij}) = (\theta_1 + \eta_j) + \sum_k \theta_{Ccov,k} \cdot (\mathrm{Ccov}_{ijk} - \mathrm{Ccov}_{med,k}) + \sum_l \theta_{Bcov,l} \cdot \mathrm{Bcov}_{ijl} + \varepsilon_{ij} \cdot \mathrm{SE}_{ij};$$

where $\theta_{Ccov,k}$ represents the rate of change in response relative to the k^{th} median-centered continuous covariate (Ccov); $\mathrm{Ccov}_{med,k}$ is the median of Ccov_k and $\theta_{Bcov,l}$ represents the shift in response for populations where the l^{th} binary covariate (Bcov) has a value of one.

Covariates were introduced using a stepwise forward selection algorithm with a likelihood ratio test based on ΔMOF. The significance level (α) chosen for covariate entry into the model was 0.05 (ΔMOF of 3.84). All available covariates were tested in the forward selection process, including but not limited to, indicators for treatment characteristics (e.g., antiangiogenic agent, class of agent, concomitant LOC therapy, and CTx) and trial characteristics (e.g., blinded, randomized, and phase II or later); percent of population with a specific characteristic (e.g., Child-Pugh B score, HBV, and male), and other characteristics (e.g., publication year or median age). Antiangiogenic agents tested were: brivanib, vandetanib, bevacizumab, erlotinib, lapatinib, sunitinib, sorafenib, and eight agents grouped as "other" (cabozantinib, cetuximab, cediranib, PTK787/ZK222584, imatinib, linifanib, tivantinib, and TSU-68).

After the forward selection procedure, the resulting model was subjected to a backward elimination algorithm using a significance level of $\alpha = 0.01$ (ΔMOF = 6.63). This process was repeated until all the remaining covariates, when excluded one at a time, resulted in significant likelihood ratio tests ($P < 0.01$). This model was considered the final model.

Because the residual error term was scaled by the observed SE of the response (ε·SE), a value of one was considered the lower limit for the estimated variance of ε (σ^2). An estimated value of σ less than one would indicate

that the model exhibited a lower variability in OS compared to what was observed in the actual study arms. In this situation, the σ term was fixed to one, and the final parameters were re-estimated using NONMEM.

Model evaluation

The final model was assessed for adequacy using both residual- and simulation-based techniques. Residual-based model diagnostics consisted of graphs of parameters derived from model fitting and empirical Bayesian estimates. Simulation-based diagnostics included assessments of parameter distributions derived from multiple simulations of the original dataset.

Residual-based diagnostics consisted of the following: observed response vs. predicted treatment arm (individual) response and observed response vs. the predicted central tendency (population) response; individual WRES vs. predicted individual response; population WRES vs. population predicted response; and histograms and quantile-quantile plots of individual WRES, population WRES, and random effects (η). Publication bias was assessed via funnel plots of model residual values and with the metabias() function within the "meta" library of the R programming language.

The final model was used to simulate the original analysis dataset 10,000 times, including uncertainty in parameter estimates (calculated from the Hessian matrix). Simulation-based diagnostics were derived from these simulated datasets. The median and 2.5th and 97.5th percentiles of each simulated mOS value were compared to the observed mOS and 95% CI for each treatment arm. Additionally, a deterministic meta-analysis was performed on each of the 10,000 simulated datasets (rma() function in metafor R-library). This meta-analysis yielded three parameters (median and 2.5th and 97.5th percentiles) per simulated dataset for comparison with the observed deterministic meta-analysis result. Density plots were used to compare the simulated distributions of these three parameters to the respective parameters obtained using the observed data.

Clinical trial simulations

The final model was used to simulate results from a phase II trial in patients with aHCC (NCT01210495).[5] In these simulations, it was assumed that axitinib would perform at least as well as other tested AATs, as measured by mOS. This assumption was based on prior knowledge that the VEGF receptor inhibition potency for axitinib is at least equivalent to other AATs. Relevant trial characteristics in the simulations included the following: 132 patients received active therapy, 66 patients received PBO, 50% of patients were HBV-positive, all patients had received prior systemic therapy, and no patient received concomitant LOC therapy or CTx. Results were simulated for 1,000,000 trials and the distribution of mOS in each treatment arm (active and PBO) was assessed, along with the ratio of mOS values (PBO:active). The distribution of the upper 97.5th percentile (upper bound of 95% CI) of the mOS ratio was used to determine the probability that the active arm would demonstrate an improved response with statistical significance.

All simulations were repeated after removing residual variability, while retaining between-study variability and uncertainty, producing results independent of trial arm size. The distribution of mOS from these simulations was assumed to represent the "truth," or what is known based on the final model. The distribution of the "true" mOS ratio was used to assess the probability that a patient population would truly benefit from active treatment compared to PBO. In order to determine the minimum number of patients needed to achieve an 80% probability of trial, all trial simulations were repeated using various patient numbers per treatment arm.

Acknowledgments. The authors would like to acknowledge and thank Carla Hernandez and Sima Ahadieh for their help with the literature search and dataset creation.

Conflict of Interest/Disclosure. The analysis described in this paper was sponsored by Pfizer. All authors are current or former employees of and hold stock in Pfizer.

Author Contributions. M.L.Z. wrote the manuscript. M.L.Z., Y.C., Y.P., O.V., and M.A.A. designed the research. M.L.Z. performed the research. M.L.Z. and D.J.N. analyzed the data.

1. Yamaguchi, R., Yano, H., Iemura, A., Ogasawara, S., Haramaki, M. & Kojiro, M. Expression of vascular endothelial growth factor in human hepatocellular carcinoma. *Hepatology* **28**, 68–77 (1998).
2. Llovet, J.M. *et al.* Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 378–390 (2008).
3. Siegel, A.B. *et al.* Phase II trial evaluating the clinical and biologic effects of bevacizumab in unresectable hepatocellular carcinoma. *J. Clin. Oncol.* **26**, 2992–2998 (2008).
4. Li, G., Hu-Lowe, D., Christensen, J. & Pocalyko, D. Abstract 4177: Inhibition of tumor malignancy by anti-angiogenic therapies in orthotopic mouse models of hepatocellular carcinoma. *Cancer Res.* **70**(8 suppl.), 4177 (2011).
5. Axitinib for the treatment of advanced hepatocellular carcinoma. <https://clinicaltrials.gov/ct2/show/NCT01210495>. Accessed 30 March 2015.
6. Xie, B., Wang, D.H. & Spechler, S.J. Sorafenib for treatment of hepatocellular carcinoma: a systematic review. *Dig. Dis. Sci.* **57**, 1122–1129 (2012).
7. Pazo-Cid, R.A. *et al.* Novel antiangiogenic therapies against advanced hepatocellular carcinoma (HCC). *Clin. Transl. Oncol.* **14**, 564–574 (2012).
8. Kang, Y.-K. *et al.* LBA17: randomised study of axitinib (AXI) plus best supportive care (BSC) versus placebo (PBO) plus BSC in patients with advanced hepatocellular carcinoma (HCC) following prior antiangiogenic therapy. *Ann. Oncol.* **25**(suppl. 4), 17 (2014).
9. Lynch, R.M., Monfort, K. & Fortune, J.R. Range estimation: a Monte Carlo examination of variability, sample size and distributional shape. *Interstat* **May** (2004).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://www.wileyonlinelibrary.com/psp4)