ARTICLE

Comparison of Various Phase I Combination Therapy Designs in Oncology for Evaluation of Early Tumor Shrinkage Using Simulations

Jérémy Seurat[1],*, Pascal Girard[2], Kosalaram Goteti[3] and France Mentré[1]

There is still a lack of efficient designs for identifying the dose response in oncology combination therapies in early clinical trials. The concentration response relationship can be identified using the early tumor shrinkage time course, which has been shown to be a good early response marker of clinical efficacy. The performance of various designs using an exposure–tumor growth inhibition model was explored using simulations. Different combination effects of new drug M and cetuximab (reference therapy) were explored first assuming no effect of M on cetuximab (to investigate the type I error (α)), and subsequently assuming additivity or synergy between cetuximab and M. One-arm, two-arm, and four-arm designs were evaluated. In the one-arm design, 60 patients received cetuximab + M. In the two-arm design, 30 patients received cetuximab and 30 received cetuximab + M. In the four-arm design, in addition to cetuximab and cetuximab + M as standard doses, combination arms with lower doses of cetuximab were evaluated (15 patients/arm). Model-based predictions or "simulated observations" of early tumor shrinkage at week 8 (ETS8) were compared between the different arms. With the same number of individuals, the one-arm design showed better statistical power than other designs but led to strong inflation of α in case of misestimated reference for ETS8 value. The two-arm design protected against this misestimation and, with the same total number of subjects, would provide higher statistical power than a four-arm design. However, a four-arm design would be helpful for exploring more doses of cetuximab in combination with M to better understand the interaction.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Currently, there is no standard for design choice in early clinical studies of combinations treatments in immuno-oncology.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ These clinical trial simulations compared designs for the first time in the context of a fixed dose of one agent assumed to be optimal while testing different doses of another based on early tumor shrinkage.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Statistical tests performed on individual model-based early tumor shrinkage at week 8 values and the one-arm design indicate a better power than two-arm or four-arm designs, but imply strong assumptions about the historical reference value, leading to strong inflation of type I error in the case of underestimating the reference. Choosing a

two-arm or a four-arm design depends on the objective of the study: a two-arm design is preferable to a four-arm design to achieve a good statistical power, but a four-arm design allows better exploration of the combination and better dose selection.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ These simulations show that, in early-phase oncology combination studies (with early tumor shrinkage as the outcome), the choice of a one-arm design (i.e., without a comparator arm) is not appropriate unless the conditions of the clinical trial are similar to those of the reference treatment or because of ethical reasons. This implies strong assumptions about the historical value, and the trial could lead to a wrong conclusion if the historical value is no longer relevant.

Early-phase studies are critical in oncology drug development. These trials are frequently designed to evaluate the overall response as well as toxicity and pharmacokinetics. Go/no go decisions and dose recommendations for further studies are based on the results of phase I studies.

Therefore, the design choice is crucial in the early phase.[1] There is growing interest in the use of combination therapies in oncology, especially immuno-oncology.[2] The combination of several agents could have a synergistic effect and enhance the antitumor activity. However, toxicities are

frequently observed in these studies, especially when chemotherapies are involved.[3] In the past two decades, several new targeted therapies, such as monoclonal antibodies, with efficacy in cancer subpopulations and limited adverse effects compared with traditional therapies have reached the market.[4] These drugs have recently been tested in many combination trials in addition of chemotherapy/radiotherapy.[5,6] However, the designs of clinical trials investigating combinations of immune-oncology agents are often based on empiricism without clear assumptions about the drug interactions. Therefore, the design choice in combination trials could be challenging, especially when clinical information is only available for single agent given as monotherapy. Recommendations have been made concerning the design of early studies with a combination of targeted anticancer therapies.[7] The importance of the assumed pharmacodynamic drug interaction and the usefulness of biomarkers are highlighted in these recommendations. More recent recommendations insist on the importance of including drug concentrations in the analysis and understanding the pharmacokinetic/pharmacodynamic relationships.[8] The importance of including biomarker-driven objectives in phase I trials has also been shown.[9]

Among clinical trial end points widely used in oncology, overall survival is universally accepted but involves large studies with long follow-up, which is not compatible with early-phase studies. End points based on tumor assessment can be used. The Response Evaluation Criteria in Solid Tumors (RECIST) criteria, widely adopted in oncology clinical trials including on targeted therapies, consist of four categories that depend on the evolution of tumor size (TS).[10,11] However, in the field of modeling and simulation it seems preferable to use the continuous TS data when available rather than RECIST as this would result in a loss of information.[12,13] Furthermore, the link between the early tumor shrinkage (ETS) and overall survival or progression free-survival has been proposed as an early efficacy marker based on several studies.[14–16]

Nonlinear mixed effect models (NLMEM) are increasingly used to support drug decision making, especially in early-phase trials.[17] In oncology, these models can be used to capture the pharmacokinetics of antitumoral agents or biomarker kinetics.[18] NLMEM can describe tumor dynamics, and several tumor growth models are found in the literature such as Gompertz[19] or Simeoni.[20] Tumor growth inhibition (TGI) models including a drug induced decrease in TS are used for various types of cancer and antitumoral agents, including targeted therapies.[15,21] TS (or other biomarkers) can also be linked to survival by a joint model.[22] However, before conducting a trial and modeling the data, it is crucial to choose an appropriate design to obtain good precision of the parameter estimates and thus precise results. For this purpose, model-based clinical trial simulation (CTS) can be performed to evaluate and compare designs.

In simulations, we used a combination of candidate drug M given on top of cetuximab, a monoclonal antibody targeting the epidermal growth factor receptor of tumor cells[23] that has been approved for metastatic colorectal cancer with wild-type RAS and usually given in combination with

chemotherapy. Several ongoing or completed studies indicate that new combinations including cetuximab could be promising for different cancers.[24,25] The aim of this study was to compare by CTS several combination designs in the treatment of refractory colorectal cancer with different assumptions about the interaction between cetuximab and M. The performances, type I error (α), and power of several designs were compared to test the superiority of the combination treatment to cetuximab alone based on ETS at week 8 (ETS8). Another objective was to compare, by modeling and simulation of exposure–TGI, the test performances on predicted ETS8 obtained from the model to the observed ETS8.

METHODS

The different models used to perform the simulations were the model of pharmacokinetic exposure during the first 8 weeks of therapy for each drug and the TGI model. We studied three different designs (one arm, two arms, and four arms) in three different scenarios: no effect of M, additivity, and synergy. Individual ETS8 values were computed as predictions after global model fitting, observations, or true values.

Models
Cetuximab and drug M pharmacokinetic exposures. Cetuximab has been described by various pharmacokinetic models.[26,27] A significant influence of body surface area (BSA) on clearance of cetuximab has been shown.[28] Therefore, it is recommended to adjust the dose of cetuximab (Dose$_C$) on BSA.[29] We assumed a simple model with independent clearance with dose, as found in a previous study in patients with metastatic colorectal cancer.[30] Therefore, cetuximab exposure expressed as area under the curve (AUC$_C$) integrated over 2 weeks (between two administrations), in mg × week/L, was simulated as in Eq. 1.

$$AUC_{C:2\ weeks} = \frac{Dose_E(mg/m^2) \times BSA(m^2)}{CL_C(L/week) \times (BSA/1.85)^{0.75}},\quad (1)$$

where CL$_C$ is the cetuximab clearance value (CL$_C$) presented in **Table 1**, including BSA effect standardized by 1.85 (i.e., the median BSA that was expected before the clinical trial), and BSA was simulated according to a lognormal distribution with a median of 1.75 and m^2 and an interindividual variability of 10% according to derived BSA distribution in previous cetuximab studies.[31–33]

The current maximal dosing regimen for cetuximab given in combination with another treatment is 500 mg per m^2 every 2 weeks. This dosing regimen has a comparable activity and safety to the standard dosing regimen.[34] A dosing regimen of 400 and 200 mg per m^2 every 2 weeks were also considered in case of combination with drug M.

Concerning the drug M, its area under the curve (AUC$_M$) was integrated over 2 weeks (between two administrations) and simulated as in Eq. 2:

$$AUC_{M:2\ weeks} = \frac{Dose_M}{CL_M},\quad (2)$$

Table 1 Parameter values for area under the curve models of cetuximab and drug M and for the tumor growth inhibition model

Parameter (unit)	μ		ω²	
Pharmacokinetics models				
CL_C (L.week⁻¹)	3.9		0.0025	
CL_M (L.week⁻¹)	5.0		0.01	
Tumor growth inhibition model				
TS0 (mm)	100[a]		0.5[a]	
KS (week⁻¹)	0.001		—	
KD_{SoC} (week⁻¹)	0.015		1.5	
KD_C (L × mg⁻¹.week⁻²)	0.00025[a]		1.5[a]	
KD_M (L × mg⁻¹.week⁻²)	0.00025[a]		1[a]	
KR (week⁻¹)	0.2		1	
Interaction parameters	**Additivity**	**Synergy**	**Additivity**	**Synergy**
INT	0	2[a]	—	0.16[a]
INT_{50} (mg × week/L)	—	75	—	—
Correlations				
Cov(η$_{KR}$,η$_{KDC}$)		1[a] (ρ = 0.82)		
Residual error				
σ²$_{prop}$		0.023[a]		

μ are the fixed effects, ω² the variance of random effects η, and σ² the variance of residuals.
CL_C, cetuximab clearance value; CL_M, clearance of drug M; Cov, covariance; INT, interaction parameter; INT_{50}, exposure needed for 50% of interaction maximal effect; KD_C, effect of drug cetuximab; KD_M, effect of drug M; KD_{SoC}, standard of care effect; KR, constant rate of resistance development; KS, natural growth of TS; L.week⁻¹, Litre.week⁻¹; η, random effects; TS0, tumor size at baseline.
[a]Estimated parameter.

where CL_M is the clearance of drug M (CL_M) presented in **Table 1**.

Tumor size. Tumor size is defined as the sum of the longest diameters of target lesions that are simulated through a TGI model.[15] The concentration of anticancer antibodies such as cetuximab is associated with the response rate.[27] A previous TGI model was developed based on clinical data from patients affected by metastatic colorectal cancer.[28] In this analysis, the TS dynamics was described by the ordinary differential equation (Eq. 3) linking TS and both drug exposures (AUC_C and AUC_M).

$$\frac{dTS}{dt} = KS \times TS - KD \times e^{-KR \times t} \times TS$$
$$KD = KD_{SoC} + KD_C \times AUC_C + KD_M \times AUC_M \times \left(1 + \frac{INT \times AUC_C}{INT_{50} + AUC_C}\right) \quad (3)$$
$$TS(0) = TS0$$

In this model (see Supplementary file), KS is the natural growth of TS, KD includes the different pooled drug effects on tumor shrinkage, KR represents the constant rate of resistance development, and TS0 is the TS at baseline. KD is the sum of KD_{SoC} (for standard of care, chemotherapy), $KD_C \times AUC_C$ (for cetuximab, with KD_C the effect of cetuximab), and $KD_M \times AUC_M \times \left(1 + \frac{INT \times AUC_C}{INT_{50} + AUC_C}\right)$ (for drug M, with KD_M the effect of M).

According to Eq. 3, the effect of drug M is the sum of its proper effect and its interaction with cetuximab. This interaction model is inspired by the global pharmacodynamics interaction model.[35] Depending on the sign of the interaction parameter (INT), this model is able to capture different types of interactions. A positive INT represents synergism, a negative INT represents antagonism, and if this parameter is null, the effects of both drugs are additive. Parameter

values of this model are given in **Table 1** based on previous reported analysis on cetuximab (excepted KD_M and interaction parameters) in the treatment of colorectal cancer.[28] The standard of care effect (KD_{SoC}) includes the placebo response. KD_M is selected from a sensitivity analysis to obtain plausible tumor shrinkage. Two interaction scenarios are investigated: additivity in which the two compounds act independently and synergy in which the cetuximab exposure increases the drug M effect. The difference between these two scenarios in terms of median TS evolution from 0 to 8 weeks is illustrated in **Figure S1**.

Let *i* denote the *i*th individual (*i* = 1,…, *N*) and *j* the *j*th TS measurement of an individual. The statistical model for the TS observation TS$_{ij}$ in individual *i* at time t_{ij} is given by Eq. 4.

$$TS_{ij} = f(t_{ij}, \theta_i) + \varepsilon_{ij}, \quad (4)$$

where *f* is the TGI model described in Eq. 3, θ_i is the vector of individual parameters, and ε_{ij} is the residual errors of mean 0 and of variance proportional to the TS: $Var(\varepsilon_{ij}) = f(t_{ij}, \theta_i)^2 \times \sigma^2_{prop}$. The random effect model is exponential as $\theta_i = \mu \times exp(\eta_i)$, where μ is the fixed effect and η_i the random effect, except for INT in which the random effect model is additive: $\theta_i = \mu + \eta_i$. The AUCs of both cetuximab and M are simulated assuming a constant clearance and used as regressors in the TGI model.

Simulation settings
Designs. We studied three types of design—one arm, two arms, or four arms—considering the same total number of individuals (*N* = 60). Simulated TS are performed at predose (baseline) and weeks 2, 4, 6, and 8. The inclusion criterion was a baseline TS > 20 mm. Dropout, incomplete adherence, missing TS measurements, and protocol

deviations were not considered. The therapy for all patients included a standard of care, which is omitted next for simplification. Simulations were performed using R software (R Foundation for Statistical Computing, Vienna, Austria) version 3.4.3., specifically, packages *mvtnorm* to generate individual parameters[36] and *deSolve* to simulate longitudinal data.[37] Simulated TS < 10 mm were censored and considered to be below the limit of quantification. The dosing regimens of cetuximab and drug M for each possible design are described in the following paragraphs. The dose of M was assumed to be fixed. For each design and each interaction scenario, 500 data sets were simulated.

One-arm: M + C500. In the one-arm design, the $N = 60$ patients were allocated to a combination M + C500: drug M and cetuximab. The dosing regimen was fixed for each drug: cetuximab at 500 mg/m^2 every 2 weeks (C500) and drug M at its standard dose.

Two-arm: C500/M + C500. In the two-arm design, the $N = 60$ patients were randomized among the two possible arms: cetuximab "alone" (C500) or M + C500, resulting in 30 patients in each arm. Typical simulated data for patients in the two arms for the additivity and synergy scenarios are presented in **Figure S2**.

Four-arm: C500/M ± C200/M ± C400/M + C500. In the four-arm design, two combination arms with lower doses of cetuximab were considered (M + C400 and M + C200) in addition to C500 and M + C500. In the lower dose arms, the dose of C was 400 mg/m^2 (C400) or 200 mg/m^2 every 2 weeks. Each arm included 15 patients.

Scenarios of the effect of M + C. Three scenarios concerning the effect of M and the interaction between M and C were considered. To investigate the type I error of tests (superiority of M + cetuximab vs. cetuximab alone), a first scenario without an effect of drug M (i.e., with $KD_M = 0$) was studied. In the other two scenarios, an effect of M ($\mu_{KDM} = 0.00025$) was implemented. These two scenarios differ in the interaction between M and cetuximab: additivity (INT = 0) or synergy ($\mu_{INT} = 2$).

Analysis
Predictions, observations, and true individual ETS8 values. For each individual i, ETS8 is expressed as a percentage and calculated as in Eq. 5.

$$ETS8_i = \frac{(TS_i(0) - TS_i(8))}{TS_i(0)} \times 100, \qquad (5)$$

where $TS_i(0)$ and $TS_i(8)$ are individual TS at baseline and after 8 weeks of treatment, respectively. Predicted (IPRED), observed (OBS), and true (TRUE) ETS8 are calculated using predictions, observations, and true individual TS at both week 0 and week 8, respectively, as described in the next paragraphs.

Individual predictions (IPRED) were obtained from model-based parameters. Parameters with a superscript letter in **Table 1** are estimated based on the TGI model (Eq. 1) by fitting the simulated data using the SAEM (Stochastic Approximation Expectation-Maximization) algorithm[38] in MONOLIX 2018R2 software (Lixoft, Antony, France). MONOLIX runs were performed with five Markov chains, a minimum of 500 iterations during the exploratory phase and a minimum of 200 iterations during the smoothing phase. An adequate method is used to account for below the limit of quantification data.[39] TS individual predictions at baseline and week 8 were obtained using the mode of empirical Bayes estimates of the individual parameters. In the synergy model, the μ_{INT} and its interindividual variability ω_{INT} were estimated, whereas in the additivity model they were fixed to 0. Simulated data without an effect of M were refitted for both the additivity and synergy models to evaluate the type I error for each assumed interaction.

Individual simulated TS observations (OBS) were directly calculated from the simulated TS values plus residual error while true individual TSs (TRUE) were simply derived from individual parameters without residual error. It is important to realize that true values are not observed in a real clinical trial. An example of OBS, IPRED, and TRUE TSs in one individual is illustrated in **Figure S3**.

Design evaluation. The performances of each design were evaluated and compared in terms of type I error and power to detect the superiority of the combination (drug M + cetuximab) to cetuximab based on the ETS8 values (OBS, IPRED, and TRUE). The statistical tests described in the next paragraphs were two-sided using α = 5%.

For the one-arm design, as all the individuals were allocated to the combination arm, individual ETS8 values were compared with a reference ETS8 for cetuximab alone from historical data using a one-sample Wilcoxon test. Different scenarios were investigated concerning the reference in which the historical value is either not valid (lower or higher than true, due, e.g., to standard of care changes or differences in patient cohorts) or true for this reference ETS8. In this study, the reference value was 29%, (i.e., the median of all individual simulated ETS8 values with C500).

For the randomized trial with two arms, C500 and cetuximab + M, a two-sample Wilcoxon test was performed to compare the ETS8 values of individuals from each arm.

In the four-arm design, a global Kruskal–Wallis test was performed first to compare the ETS8 values between the four arms. If a difference between at least two arms was detected by this global test, a Dunnett test was performed to compare each combination arm to the reference C500 alone using a Benjamini–Hochberg procedure to account for multiplicity of tests.[40]

From the data sets simulated without an effect of drug M ($KD_M = 0$, i.e., under H$_0$), the type I error was evaluated as the proportion of trials for which H$_0$ is rejected and compared to the nominal value of 5%. The 95% prediction interval for proportion π of H$_0$ rejection (type I error or power) is given by binomial law[41] ((3.3%, 7.3%) with 500 data sets). From the data sets simulated with an effect of drug M (additivity or synergy scenario, i.e., under H$_1$), the power of the tests was evaluated as the proportion of trials for which H$_0$ is rejected.

RESULTS

First, we checked whether model-based ETS8 values (IPRED) were overpredicting or underpredicting the true values. Overall, ETS8 are distributed accurately around the identity line for each design and under both additivity and synergy scenarios. However, a trend of overpredicted values for ETS8 < 25% is shown, although it seems to not be influenced by the scenario, design, or treatment arm (**Figure S4**).

One-arm: M + C500

For the one-arm design, the ETS8 values for patients were compared to an ETS8 value for current reference therapy obtained from historical data. Distributions of individual ETS8 (OBS, IPRED, and TRUE) are presented in **Figure 1** in the cases of additivity and synergy. The median of OBS, IPRED, and TRUE ETS8 values were close to each other in each scenario. The observed median ETS8 was 43% for additivity and 55% for synergy, whereas the reference ETS8 for cetuximab alone was 29%. According to our model and simulation settings, assuming additivity, almost three of four patients had greater tumor shrinkage at week 8 than those of reference treatment. These boxplots enable us to see the interindividual variability in ETS8. As predictions account for residual error, the 5th to 95th percentile of ETS8 is narrower for IPRED (from 23% to 78% for additivity and from 28% to 92% for synergy) than OBS (from 3% to 78% for additivity and from 13% to 89% for synergy). In both scenarios, the distribution of predictions seems to be more representative of true values than observations.

With this one-arm design, as the ETS8 distribution is only known for M + C500, a one-sample Wilcoxon test was performed (**Table 2**). Using observed data and assuming a reliable ETS8 reference value under C500 (in our case 29%), the type I error is controlled (5.6%) as it lies in the prediction interval around the nominal value α = 5%. However, the type I error is not controlled using model-based predictions (12.6% for additivity and 13.2% for synergy), which is partly attributed to a lack of identifiability between the M and cetuximab effect. Overall, assuming a reliable reference, good power (at least 89%) is reached using either OBS or IPRED.

In this kind of design, for multiple possible reasons, the assumed reference value from historical data could be wrong. Therefore, we investigated the type I error and power with lower (25% and 20%) and higher (35% and 40%) ETS8 reference values. Performing a two-sided Wilcoxon test, the type I error was inflated in cases using either OBS or TRUE. For example, if we compare the ETS8 distribution to 20% instead of 29%, we wrongly conclude to superiority of the combination in more than the half of the CTS using observations or predictions. Moreover, assuming a higher reference value than true leads to a lack of power. For example, in the additivity scenario, assuming a reference value of 35% leads to a power of 45.2% using observations (instead of 89.0% with the real reference value).

Two-arms: C500/M + C500

For the randomized two-arm design, the distribution of individual ETS8 (OBS, IPRED, and TRUE) values in each arm is shown in **Figure 2** for additivity and synergy. As with the one-arm design, the median OBS, IPRED, and TRUE ETS8 values are close to each other in each arm and each scenario. For C500, the median ETS8 is ~ 29% (i.e., the reference value in the one-arm design) and for M + C500, the median ETS8 is ~ 43% with additivity and ~ 55% with synergy. Notably, with C500, more than 1 of 10 patients presented a negative observed ETS8 (i.e., increasing TS between weeks 0 and 8), whereas none of the individuals had a true negative ETS8 under this therapy with our simulation settings, which is captured well using predicted ETS8. The IPRED distribution of C500 seems to be similar between the two scenarios. This indicates that the estimated cetuximab specific effect is not influenced by the model or the scenario, as in the simulated data, and allows for comparison between the scenarios. As with the individuals given M + C500 in the one-arm design, the variability in ETS8 is considerable with C500 in the two-arm designs, especially in OBS (the 5th to 95th percentile extends from –12% to 65%). Therefore, the difference in ETS8 distribution between the two arms is difficult to distinguish in the case of additivity. The IPRED distribution is narrower and closer to TRUE distribution than OBS.

With this design, comparing ETS8 distributions between the two arms by performing two-sample Wilcoxon tests, no assumption on ETS8 reference value is required (**Table 3**). Therefore, choosing a two-arm design instead of a one-arm design avoids a potential inflation of the type I error. In both
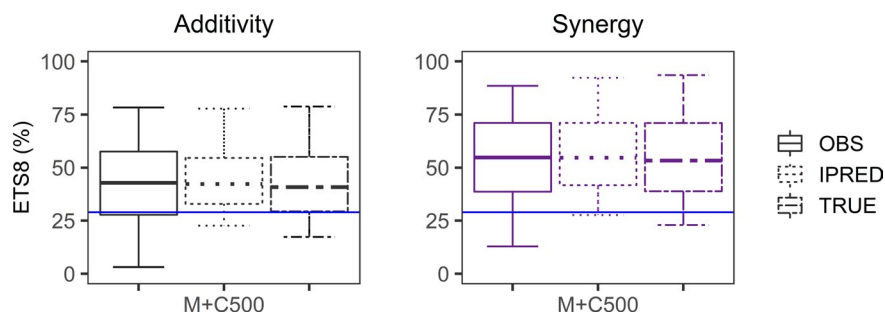

Figure 1 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the one-arm design (combination M + C500). The blue horizontal line is the true median ETS8 (29%) of the historical reference treatment. Whiskers indicate 5th to 95th percentiles.

Table 2 Type I error and power of tests in one-arm design

Theoretical median of ETS8 from historical data (%)		Additivity						Synergy					
		Type I error (%)			Power (%)			Type I error (%)			Power (%)		
		OBS	IPRED	TRUE	OBS	IPRED	TRUE	OBS	IPRED	TRUE	OBS	IPRED	TRUE
Lower than true	20	56.6	79.4	68.6	100.0	100.0	100.0	56.6	60.2	68.6	100.0	100.0	100.0
	25	13.2	34.2	18.0	98.6	99.6	99.8	13.2	34.2	18.0	100.0	100.0	100.0
True reference	29	5.6	12.6	6.0	89.0	97.6	93.6	5.6	13.2	6.0	100.0	100.0	100.0
Higher than true	35	27.4	31.6	37.4	45.2	70.6	51.0	27.4	32.2	37.4	99.4	100.0	99.8
	40	74.8	75.0	82.6	9.8	25.0	10.6	74.8	74.8	82.6	90.2	98.0	93.0

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. Under H_0, the 95% prediction interval from binomial distribution is 3.3%–7.3% for 500 simulations. For the one-arm design, the number of individuals by data set is $N = 60$, and all individuals are allocated to the combination arm. The ETS8 distribution (of observations (OBS), predictions (IPRED), or true values (TRUE)) is compared with a theoretical median of ETS8 under cetuximab alone, performing a one-sample Wilcoxon test (two-sided, $\alpha = 5\%$). Different possible theoretical median of ETS8 values (true or not) are investigated. Two possible drug interactions are considered: additivity and synergy (see Models section).
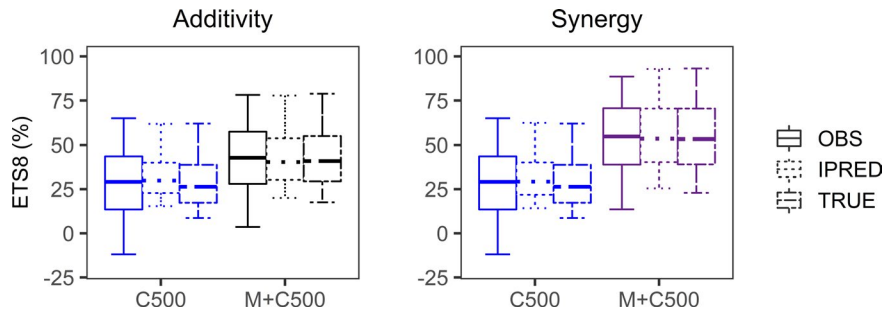

Figure 2 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the two-arm design. Whiskers indicate 5th to 95th percentiles. C500 is cetuximab alone vs. the combination treatment (M + C500).

additivity and synergy scenarios and using either observed or predicted ETS8, the type I error is well controlled as it lies in the 95% prediction interval (3.3%–7.3%). Compared with the previous design, there is a decrease in power in the additivity scenario (using OBS: 66.6% vs. 89.0% in one-arm design). This result is explained by having only 30 patients per arm (vs. 60 in the one-arm design). Nevertheless, this design remains powerful in the case of synergy (using OBS: 97.4% vs. 100% in the one-arm design). Moreover, in each scenario, using the model-based ETS8 value leads to a slight increase in power compared with using the raw data. For example, the power in the additivity scenario was 69.2% using IPRED.

Four-arms: C500/M + C500/M + C400/M + C200
For the four-arm design, the distributions of individual ETS8 values (OBS, IPRED, and TRUE) in each of the arms are presented in **Figure 3** for additivity and synergy. The IPRED distribution seems to be closer to TRUE than OBS, especially for extreme percentiles, except in the C500 arm. As expected, the distributions of C500 and M + C200 are close to each other for additivity. Moreover, in synergy, the ETS8 distribution is close between the two combination arms with the higher doses of cetuximab.

Type I error and power are presented in **Table 4**. Performing a global Kruskal–Wallis to detect a difference between the ETS8 distributions of the four arms, more power was

Table 3 Type I error and power of tests in two-arm design

	Additivity		Synergy	
	Type I error (%)	Power (%)	Type I error (%)	Power (%)
OBS	6.2	66.6	6.2	97.4
IPRED	6.2	69.2	7.0	98.6
TRUE	6.2	89.0	6.2	99.8

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. Under H_0, the 95% prediction interval from binomial distribution is 3.3%–7.3% for 500 simulations. For the two-arms design, 30 patients are allocated to the combination arm, and 30 are allocated to cetuximab alone. The ETS8 distributions (of observations (OBS), predictions (IPRED), or true values (TRUE)) are compared between the two arms, performing a two-sample Wilcoxon test (two-sided, $\alpha = 5\%$). Two possible drug interactions are considered: additivity and synergy (see Models section).
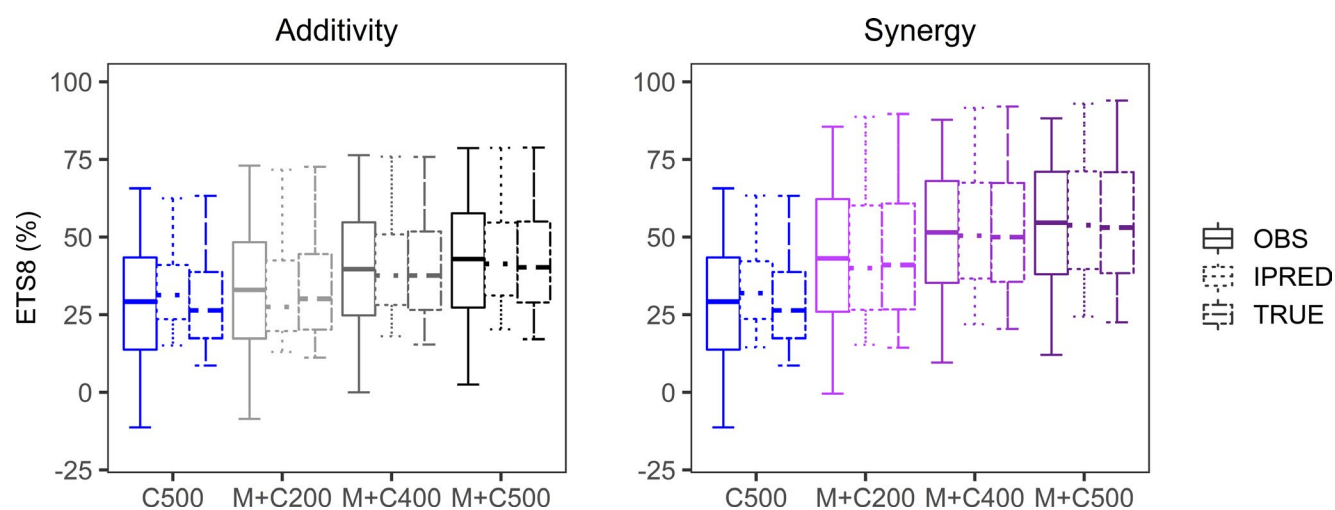
Figure 3 Boxplots of early tumor shrinkage at week 8 (ETS8) from observations (OBS; solid boxes, left subgroup for each treatment), predictions (IPRED; dotted boxes, middle subgroup for each treatment), and true values (TRUE; dashed boxes, right subgroup for each treatment) in the four-arm design. Whiskers indicate 5th to 95th percentiles. C500 is cetuximab alone, and M + C200, M + C400, and M + C500 are combination treatments.

obtained in the synergy scenario than the additivity scenario. For example, using observed ETS8, the power was 69.4% in synergy vs. 30.2% in additivity. If the global test was significant, a Dunnett test was performed, comparing the ETS8 distribution in each of the three combination arms to those of the reference arm C500. Under the null hypothesis, if we compare the distribution of ETS8 for C500 and M + C500, the Dunnett test was conservative compared with a single Wilcoxon test, the type I error using observations was 2% with this procedure in the four-arm design vs. 6.2% for the Wilcoxon test in the two-arm design. Therefore, in the additivity scenario, the power to detect the superiority of M + C500 vs. C500 decreased to 19.8% using observations and 25.6%

using predictions (vs. 66.6% and 69.2% in two-arm design, respectively), which is partly explained by the lesser number of patients per arm (15 vs. 30). Nevertheless, in synergy and using model-based ETS8, we achieved 65.4% power to detect the superiority for M + C500 and 52% for M + C400. Thus, under certain conditions, this design is interesting for exploring other doses in the combination of two drugs.

DISCUSSION

This work compares designs for early clinical combination trials in oncology in the context of a fixed dose of one agent and different doses of another with the same number of

Table 4 Type I error and power of tests in four-arm design

	Additivity	Synergy
Power of Kruskal–Wallis test (%)		
OBS	30.2	69.4
IPRED	52.0	76.6
TRUE	44.0	87.8
Type I error of Dunnett test (%) C500 vs. M + C500		
OBS	1.6	1.6
IPRED	2.0	2.0
TRUE	2.4	2.4

	C500 vs. M + C200	C500 vs. M + C400	C500 vs. M + C500	C500 vs. M + C200	C500 vs. M + C400	C500 vs. M + C500
Power of Dunnett test (%)						
OBS	3.0	13.0	19.8	26.4	51.8	60.2
IPRED	5.4	13.2	25.6	18.6	52.0	65.4
TRUE	4.0	19.8	29.2	34.0	70.4	79.0

The type I error and power from clinical trial simulations are given as the observed proportion of significant tests performed on early tumor shrinkage at week 8 (ETS8) from 500 simulated clinical trials. For the four-arms design, 15 patients are allocated to each arm. The ETS8 distributions (of observations (OBS), predictions (IPRED), or true values (TRUE)) are compared between the arms, performing first a global Kruskal–Wallis test (two-sided, α = 5%), then, if the global test is significant, a Dunnett test where distribution of each combination arm (M + C200, M + C400, and M + C500) is compared with those of cetuximab alone (C500). Two possible drug interactions are considered: additivity and synergy (see Models section).

subjects and based on ETS. The one-arm design leads to inflation of type I error in the case of a misestimated reference value, as reported previously.[42] Therefore, before the analysis of data from a one-arm design, we should ensure that the conditions of the study are similar to the reference study. With the same number of subjects and an accurate reference value, the power of a one-arm design is higher than that of a two-arm design. However, to avoid the risk of an inaccurate reference value, a two-arm randomized design including an arm with the reference therapy is preferable. Such a design conducts to control the type I error. Another possibility is to choose a four-arm design with different possible dosing regimens. Although with the same total number of patients a two-arm design leads to a greater power of tests, a four-arm design offers more accurate dose selection.

Our work aimed to show simulations in an ideal setting in order to facilitate the interpretation of the results. In these simulations, we did not account for possible incomplete adherence, missing data, protocol deviations (e.g., timing of TS measurements), or dropout. It would be interesting to incorporate these design departures in future simulations and analyses to mimic a real clinical study.[43] Nevertheless, most of these departures are not preponderant in the first 8 weeks of a trial and therefore should not change drastically our simulation results. In this study, exposure to cetuximab was simulated from a previous pharmacokinetics model[28] that included a BSA effect on clearance but other covariates, such as normal fat mass, would be more physiological.[44] Moreover, the allometric coefficient of 0.75 (used for the BSA effect in the model) would be more appropriate with the normal fat mass. In our study, the clearance of M was not influenced by any covariate, but the effect of body weight or the normal fat mass could have been considered. In the model that we used, we assumed linear effects in the exposure ranges of cetuximab and M based on previous models in which the maximal effect (Emax) could not be estimated. However, it would be relevant to explore more mechanistic models and assume nonlinear effects, as is the case for the interaction between cetuximab and M in the model. The TGI model involves a global resistance parameter KR. It would be relevant to associate different resistance parameters with each drug and its exposure.

Using NLMEM to analyze longitudinal data is a promising tool, especially when the outcome is based on TS. Tumor measurements are notably influenced by scan variability and the individual reviewing the image, which may result in noise in the data[45] and a loss of power. The NLMEM accounts for the information from the longitudinal data, which were five TS measurements in our study, and allows estimation of the residual error, including the noise in the data. Therefore, performing tests based on model-based predictions can lead to better power than observations. For example, in our two-arm design with the additivity scenario, the power using model-based ETS8 was 69.2% vs. 66.6% using observations.

In this study, ETS was chosen as outcome. We could also dichotomize ETS as responder or nonresponder outcome. Different thresholds can be chosen to separate the patients depending on their ETS.[46] We can dichotomize the

subjects as responders if ETS8 ≥ 30% or nonresponder if ETS8 < 30%, which corresponds roughly to a complete or partial response vs. stable or progressive disease according to the RECIST criteria. With this binary outcome, in our two-arm design and additivity scenario, the power of Fisher's exact test was 46.4% instead of 69.2% with the Wilcoxon test of continuous model-based ETS8 values.

Guidelines highlight the importance of design choice for an early-phase combination trial in oncology and that assumptions and knowledge about pharmacokinetic/pharmacodynamic interactions should be taken into account to select the most effective trial design.[47] There is no standard design for combination studies, and designs that include randomization helps to define the effect of the combination vs. one or both of the individual components.[1]

We plan to extend this work to perform model-based adaptive two-stage designs[48,49] using the Fisher information matrix to optimize the second stage of the study[50] in which different dose arms could also be added or dropped at the end of the first stage.

Conflict of Interest. Pascal Girard and Kosalaram Goteti are employed by Merck Institute for Pharmacometrics and EMD Serono R&D Institute, Inc., respectively, an affiliate of Merck KGaA, Darmstadt, Germany. All other authors declared no competing interests for this work.

Author Contributions. J.S., P.G., K.G., and F.M. wrote the manuscript. J.S., P.G., K.G., and F.M. designed the research. J.S., P.G., K.G., and F.M. performed the research. J.S., P.G., K.G., and F.M. analyzed the data.

Disclaimer. As Editor-in-Chief of *CPT: Pharmacometrics and Systems Pharmacology*, France Mentré was not involved in the review or decision process for this article.

1. LoRusso, P.M., Boerner, S.A. & Seymour, L. An overview of the optimal planning, design, and conduct of phase I studies of new therapeutics. *Clin. Cancer Res.* **16**, 1710–1718 (2010).
2. Ledford, H. Cocktails for cancer with a measure of immunotherapy. *Nature* **532**, 162–164 (2016).
3. Sandler, A. *et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N. Engl. J. Med.* **355**, 2542–2550 (2006).
4. Huang, J. *et al.* Emerging trends in us oncological approvals: a 13-year review (1999–2011). *Drug Inf. J.* **46**, 344–357 (2012).
5. Cunningham, D. *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **351**, 337–345 (2004).
6. Sullivan, R.J. *et al.* Atezolizumab plus cobimetinib and vemurafenib in BRAF-mutated melanoma patients. *Nat. Med.* **25**, 929–935 (2019).
7. Seymour, L.K. *et al.* Design and conduct of early clinical studies of two or more targeted anticancer therapies: recommendations from the task force on methodology

for the development of innovative cancer therapies. *Eur. J. Cancer Oxf. Engl. 1990* **49**, 1808–1814 (2013).

8. Smoragiewicz, M. *et al.* Design and conduct of early clinical studies of immunotherapy: recommendations from the task force on Methodology for the Development of Innovative Cancer Therapies 2019 (MDICT). *Clin. Cancer Res.* **26**, 2461–2465 (2020).

9. Paller, C.J. *et al.* Factors affecting combination trial success (FACTS): investigator survey results on early-phase combination trials. *Front. Med.* **6** (2019). https://doi.org/10.3389/fmed.2019.00122.

10. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).

11. Litière, S. *et al.* RECIST 1.1 for response evaluation apply not only to chemotherapy-treated patients but also to targeted cancer agents: a pooled database analysis. *J. Clin. Oncol.* **37**, 1102–1110 (2019).

12. Sharma, M.R., Maitland, M.L. & Ratain, M.J. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox–point. *Cancer Res.* **72**, 5145–5149 (2012), discussion 5150.

13. Vera-Yunca, D. *et al.* Machine learning analysis of individual tumor lesions in four metastatic colorectal cancer clinical studies: linking tumor heterogeneity to overall survival. *AAPS J.* **22**, 58 (2020).

14. 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).

15. 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).

16. Colloca, G.A., Venturino, A. & Guarneri, D. Early tumor shrinkage after first-line medical treatment of metastatic colorectal cancer: a meta-analysis. *Int. J. Clin. Oncol.* **24**, 231–240 (2019).

17. Kimko, H. & Pinheiro, J. Model-based clinical drug development in the past, present and future: a commentary. *Br. J. Clin. Pharmacol.* **79**, 108–116 (2015).

18. Petrylak, D.P. *et al.* Evaluation of prostate-specific antigen declines for surrogacy in patients treated on SWOG 99–16. *J. Natl. Cancer Inst.* **98**, 516–521 (2006).

19. Gompertz, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. R. Soc. Lond.* **115**, 513–583 (1825).

20. Simeoni, M. *et al.* Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res.* **64**, 1094–1101 (2004).

21. Bonate, P.L. & Suttle, A.B. Modeling tumor growth kinetics after treatment with pazopanib or placebo in patients with renal cell carcinoma. *Cancer Chemother. Pharmacol.* **72**, 231–240 (2013).

22. Tardivon, C. *et al.* Association between tumor size kinetics and survival in patients with urothelial carcinoma treated with atezolizumab: implication for patient follow-up. *Clin. Pharmacol. Ther.* **106**, 810–820 (2019).

23. Jonker, D.J. *et al.* Cetuximab for the treatment of colorectal cancer. *N. Engl. J. Med.* **357**, 2040–2048 (2007).

24. Garrett, C.R. *et al.* Phase I dose-escalation study to determine the safety, pharmacokinetics and pharmacodynamics of brivanib alaninate in combination with full-dose cetuximab in patients with advanced gastrointestinal malignancies who have failed prior therapy. *Br. J. Cancer* **105**, 44–52 (2011).

25. Van Cutsem, E. *et al.* Binimetinib, encorafenib, and cetuximab triplet therapy for patients with BRAF V600E-mutant metastatic colorectal cancer: safety lead-in results from the phase III beacon colorectal cancer study. *J. Clin. Oncol.* **37**, 1460–1469 (2019).

26. Le Louedec, F. *et al.* Cetuximab pharmacokinetic/pharmacodynamics relationships in advanced head and neck carcinoma patients. *Br. J. Clin. Pharmacol.* **85**, 1357–1366 (2019).

27. Azzopardi, N. *et al.* Cetuximab pharmacokinetics influences progression-free survival of metastatic colorectal cancer patients. *Clin. Cancer Res.* **17**, 6329–6337 (2011).

28. Girard, P. *et al.* Drug-disease model of tumor size and overall survival in metastatic colorectal cancer patients treated with cetuximab administered weekly or every second week. European Cancer Congress, Amsterdam, Netherlands. (2013).

29. Tan, A.R. *et al.* Pharmacokinetics of cetuximab after administration of escalating single dosing and weekly fixed dosing in patients with solid tumors. *Clin. Cancer Res.* **12**, 6517–6522 (2006).

30. Tabernero, J. *et al.* Cetuximab administered once every second week to patients with metastatic colorectal cancer: a two-part pharmacokinetic/pharmacodynamic phase I dose-escalation study. *Ann. Oncol.* **21**, 1537–1545 (2010).

31. Bokemeyer, C. *et al.* Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J. Clin. Oncol.* **27**, 663–671 (2009).

32. Van Cutsem, E. *et al.* Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 1408–1417 (2009).

33. Cheng, A.-L. *et al.* Efficacy, tolerability, and biomarker analyses of once-every-2-weeks cetuximab plus first-line FOLFOX or FOLFIRI in patients with KRAS or all RAS wild-type metastatic colorectal cancer: the phase 2 APEC study. *Clin. Colorectal Cancer* **16**, e73–e88 (2017).

34. Fernandez-Plana, J. *et al.* Biweekly cetuximab in combination with FOLFOX-4 in the first-line treatment of wild-type KRAS metastatic colorectal cancer: final results of a phase II, open-label, clinical trial (OPTIMIX-ACROSS Study). *BMC Cancer* **14**, 865 (2014).

35. Wicha, S.G., Chen, C., Clewe, O. & Simonsson, U.S.H. A general pharmacodynamic interaction model identifies perpetrators and victims in drug interactions. *Nat. Commun.* **8**, 2129 (2017).

36. Genz, A. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Stat.* **1**, 141–149 (1992).

37. Soetaert, K., Petzoldt, T. & Setzer, R.W. Solving differential equations in *R*: package desolve. *J. Stat. Softw.* **33**, 9 (2010).

38. Kuhn, E. & Lavielle, M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Stat. Data Anal.* **49**, 1020–1038 (2005).

39. Ahn, J.E., Karlsson, M.O., Dunne, A. & Ludden, T.M. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *J. Pharmacokinet. Pharmacodyn.* **35**, 401–421 (2008).

40. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

41. Clopper, C.J. & Pearson, E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413 (1934).

42. Viele, K. *et al.* Use of historical control data for assessing treatment effects in clinical trials. *Pharm. Stat.* **13**, 41–54 (2014).

43. Sheiner, L.B. Is intent-to-treat analysis always (ever) enough? *Br. J. Clin. Pharmacol.* **54**, 203–211 (2002).

44. Holford, N.H.G. & Anderson, B.J. Allometric size: The scientific theory and extension to normal fat mass. *Eur. J. Pharm. Sci.* **109S**, S59–S64 (2017).

45. Sullivan, D.C., Schwartz, L.H. & Zhao, B. The imaging viewpoint: how imaging affects determination of progression-free survival. *Clin. Cancer Res.* **19**, 2621–2628 (2013).

46. Mandrekar, S.J. *et al.* Evaluation of alternate categorical tumor metrics and cut points for response categorization using the RECIST 1.1 data warehouse. *J. Clin. Oncol.* **32**, 841–850 (2014).

47. Paller, C.J. *et al.* Design of phase I combination trials: recommendations of the clinical trial design task force of the nci investigational drug steering committee. *Clin. Cancer Res.* **20**, 4210–4217 (2014).

48. Foo, L.K. & Duffull, S. Adaptive optimal design for bridging studies with an application to population pharmacokinetic studies. *Pharm. Res.* **29**, 1530–1543 (2012).

49. Pierrillas, P.B. *et al.* Model-based adaptive optimal design (MBAOD) improves combination dose finding designs: an example in oncology. *AAPS J.* **20**, 39 (2018).

50. Mentré, F., Mallet, A. & Baccar, D. Optimal design in random-effects regression models. *Biometrika* **84**, 429–442 (1997).