

# Tumor Size and Overall Survival in Patients With Platinum-Resistant Ovarian Cancer Treated With Chemotherapy and Bevacizumab

Alexandre Sostelly<sup>1</sup> and François Mercier<sup>1</sup>

Clinical Pharmacology and Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland.

Clinical Medicine Insights: Oncology  
Volume 13: 1–10  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1179554919852071



## ABSTRACT

**INTRODUCTION:** Ovarian cancer is now recognized as a constellation of distinct subtypes of neoplasia involving the ovary and related structures. As a consequence of this heterogeneity, the analysis of covariates influencing the overall survival is crucial in this disease segment. In this work, an overall survival model incorporating tumor kinetics metrics in patients with platinum-resistant ovarian cancer was developed from the randomized, open label, phase 3 AURELIA trial.

**METHODS:** Tumor size data from 361 patients randomly allocated to the bevacizumab + chemotherapy or chemotherapy study arm were collected at baseline and every 8 to 9 weeks until disease progression. Patients continued to be followed for survival after treatment discontinuation. A landmarked Cox proportional hazard survival model was developed to characterize the overall survival distribution.

**RESULTS:** Two sets of factors were found to be influential on survival time: those describing the type and severity of disease (Eastern Cooperative Oncology Group [ECOG], Fédération Internationale de Gynécologie et d'Obstétrique [FIGO] stages, presence of ascites) and those summarizing the key features of the tumor kinetic model (tumor shrinkage at week 8 and tumor size at treatment onset). The treatment group was not required in the final model as the drug effect was accounted for in the tumor kinetics model.

**CONCLUSIONS:** This work has identified both ascites and tumor kinetics metrics as being the 2 most influential factors to explain variability in overall survival in patients with platinum-resistant ovarian cancer.

**KEYWORDS:** ovarian cancer, angiogenesis, overall survival, tumor kinetics, anti-angiogenesis, therapy

**RECEIVED:** October 14, 2018. **ACCEPTED:** April 23, 2019.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The AURELIA trial was sponsored by F. Hoffmann-La Roche.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Alexandre Sostelly and François Mercier were employed by F. Hoffmann-La Roche Ltd.

**CORRESPONDING AUTHOR:** François Mercier, Clinical Pharmacology and Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 70, 4070 Basel, Switzerland. Email: francois.mercier@roche.com

## Introduction

Epithelial ovarian cancer is the fifth most common cause of cancer-related death in women. The combination of surgery and platinum-based chemotherapy has been widely used to control advanced ovarian cancers resulting in approximately 70% response rate.<sup>1</sup> Patients with disease relapses within 6 months after platinum-based chemotherapy are considered to have platinum-resistant ovarian cancer (PROC). The randomized, open-label, phase 3 AURELIA trial was designed to compare progression-free survival (PFS) in patients with PROC treated with chemotherapy (CT) alone or in combination of bevacizumab (B).<sup>2</sup>

Commonly, objective response rates (ORR) and/or PFS are used to evaluate clinical benefits of oncology drugs and may allow regulatory submission before overall survival (OS) data become mature. However, these statistics have shown limitations.<sup>3–9</sup> Other measures describing the change in tumor size over time (tumor kinetics [TK]) have been considered and have shown to be correlated with OS, at the individual level, in various indications including colorectal cancer,<sup>10</sup> gastrointestinal stromal tumor cancer,<sup>11</sup> renal carcinoma cancer,<sup>12,13</sup> and non-small cell lung cancer<sup>14</sup> and may offer an opportunity for

early evaluation of clinical benefits. This approach requires the development of paradigms linking changes in tumor size over time to models predicting clinical outcomes such as OS.<sup>15</sup>

As the relationship between TK and OS is specific to the target treatment population, survival models incorporating tumor kinetic metrics have to be developed for each tumor type and line of treatment.<sup>15</sup> Although a survival model was previously developed for patients with PROC,<sup>16</sup> there is no model relating change in TK to OS established in PROC. We aimed at developing a framework to quantify the benefits of bevacizumab in patients with PROC on TK and to assess the level of correlation between TK metrics and OS in this disease segment.

## Patients and Methods

### *Trials and data*

Individual data were available from the randomized phase 3 trial AURELIA. Details of patient characteristics and study design can be found in Pujade-Lauraine et al.<sup>2</sup> A total of 361 patients were randomly allocated to the B + CT (n=179) or CT (n=182) arm. Investigators' selection of CT was evenly distributed in both study arms among the 3 options: in all, 126



**Table 1.** Demographic and baseline patient characteristics.

	CT (n=182)	B + CT (n=179)
Age (median, range in years)	61.0 [25, 84]	62 [25, 80]
Origin of cancer: Ovary	86%	93%
Histology type: Serous	84%	87%
ECOG status: 0	54%	60%
Pre-treatment SLD (median, range in mm)	56.5 [10, 370]	54 [10, 314]

Abbreviations: B, bevacizumab; CT, chemotherapy; ECOG, Eastern Cooperative Oncology Group; SLD, sum of longest diameters.

patients (35%) received pegylated liposomal doxorubicin (PLD), 115 patients (32%) received paclitaxel, and 120 patients (33%) received topotecan. B was given at a dose of 10 mg/kg every 2 weeks or 15 mg/kg every 3 weeks depending on the CT schedule. Patients' characteristics at baseline are presented in Table 1.

Tumor size reported as the sum of longest diameters of target lesions (SLD) was defined according to RECIST 1.0 and was collected at baseline (up to 8 weeks before treatment initiation) and every 8 to 9 weeks thereafter until disease progression. Patients could discontinue earlier if there was evidence of progressive disease, if unacceptable toxicity occurred, or if the attending physician or the patient requested discontinuation. Yet, even in this case, patients continued to be followed for OS.

The AURELIA study demonstrated a prolonged PFS in patients treated with B + CT compared with CT. The study was not designed to conclude on OS, and the hazard ratio (HR) of 0.85 in favor of B + CT was associated with a 95% confidence interval (CI) ranging from 0.66 to 1.08.<sup>2</sup>

### Model development

**Tumor kinetics model.** A TK model was used to fit the SLD values. The TK model accounted for the dynamics of tumor growth, antitumor drug effect, and resistance to the drug effect (equation (1)). This model was initially proposed by Claret et al<sup>17</sup>

$$Y_{ij} = Y_{0i} \times \exp \left[ k_{rgi} \times t_{ij} - \frac{k_{si}}{\lambda_i} \times \left( 1 - \exp(-\lambda_i \times t_{ij}) \right) \right] + \varepsilon_{ij} \quad (1)$$

where  $Y_{ij}$  is the SLD value at time  $j$  in patient  $i$ ,  $Y_{0i}$  is the estimated tumor size at baseline (time of treatment onset,  $t=0$ ) in patient  $i$ ,  $k_{rgi}$  is the estimated SLD growth rate in patient  $i$ ,  $k_{si}$  is the estimated SLD decrease rate in patient  $i$ , and  $\lambda_i$  is the resistance parameter reflecting the loss of treatment effect over time in patient  $i$ . The additive residual error term  $\varepsilon_{ij}$  captures the unexplained variability in the data, including measurement errors.

The influence of treatment was tested in the model by introducing an interaction term

$$P \times TRT$$

with  $TRT = CT, B + CT$ , CT being topotecan, PLD or paclitaxel, and B + CT being any combinations of CT with B. The parameters  $k_i$  and  $\lambda$  were expected to be influenced by the treatment.

Only patients who had a pre-treatment and at least one on-treatment TS assessment were retained in the time series analysis. Overall, 9.5% of the observed SLD values were reported below the limit of detection (5 mm) and were set to 2.5 mm.

To account for deviations from a typical population parameter, inter-individual variability was incorporated as exponential random effects, such that

$$\exp(\widehat{P}_i) \sim N(P, \omega_p^2)$$

with  $P \in \{Y_0, k_{rg}, k_s, \lambda\}$ .

The model parameters were estimated using a nonlinear mixed effect approach in which all the data from all evaluable patients were simultaneously analyzed. The first-order conditional estimation method with interaction in NONMEM version 7.2 (Globomax, LLC, Hanover, USA) was used to estimate the model parameters.

The model associated with the lowest Akaike information criteria (AIC) was retained as the best model among those tested. Models with an unsuccessful estimation of parameter standard errors were not retained. To evaluate the predictive performances of the TK model, model-predicted distributions of SLD were overlaid to the observed ones.

Following previous works, 3 individual TK metrics were derived for each patient.<sup>10</sup> Early tumor shrinkage (*ETS8*) was derived as the ratio between the predicted tumor size at week 8 and the predicted tumor size at baseline; maximum tumor shrinkage (*MaxSh*) was calculated as the ratio between the predicted minimum tumor size (during CT ± B treatment) and the predicted tumor size at baseline; and time to (re-) growth (*TTG*) was the model predicted time to reach the predicted minimum tumor size (during CT ± B treatment).

**Overall survival model.** A landmarked Cox proportional-hazard model was used to analyze OS. It included clinically relevant baseline covariates such as the Eastern Cooperative

**Table 2.** Tumor kinetics metrics per treatment group.

	CT (n=140)	B + CT (n=134)
<i>ETS8</i> <sup>a</sup> (%), median [Q1, Q3]	-5.0 [-1.3, -21.0]	-17.9 [-9.5, -30.2]
<i>MaxSh</i> <sup>b</sup> (%), median [Q1, Q3]	-5.8 [-3.1, -26.7]	-22.0 [-9.9, -50.6]
<i>TTG</i> <sup>c</sup> (weeks), median [Q1, Q3]	4.3 [2.5, 19.2]	17.5 [8.4, 40.9]

Abbreviations: B, bevacizumab; CT, chemotherapy; *TTG*, time to (re-)growth; *SLD*, sum of longest diameters.

<sup>a</sup> *ETS8*: model-predicted percentage change between predicted baseline *SLD* value and predicted *SLD* value at week 8.

<sup>b</sup> *MaxSh*: model-predicted percentage change between predicted baseline *SLD* value and predicted minimum *SLD* value on treatment.

<sup>c</sup> *TTG*: time to reach the predicted minimum *SLD* value on treatment.

Oncology Group (ECOG) performance status, Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) stage at baseline, histological grade and subtype, time from first diagnosis to randomization, presence of ascites, model-predicted *SLD* at time of treatment onset ( $\hat{Y}_0$ ), and CA-125 at baseline which were tested as OS prognostic factors.

The landmark was set to week 8, corresponding to the time of first CT scan collection. As a consequence, any event (death or censoring) occurring between baseline and week 8 was removed from the OS analysis. In this framework, only the TK metrics *ETS8* was tested as a predictive biomarker in the Cox model.

A resampling procedure (using the `boot.stepAIC` function which builds on the `stepAIC` function of the R MASS package)<sup>18</sup> was used to reduce the risk of false selection.<sup>19</sup>

The predictive value of each covariate was evaluated using the concordance index (*c*-index) introduced by Harrell<sup>20</sup> and the Brier score. Harrell's overall *c*-index indicates the proportion of all pairs of subjects who can be ordered such that the subject with higher predicted survival is the one who actually survived longer. The Brier score corresponds to the squared differences between actual and predicted outcomes. Both *c*-index and Brier score are obtained using the `pec` package in R.<sup>21</sup>

## Results

### Tumor kinetics model

Of the 361 patients randomized in the AURELIA trial, 275 patients (140 in the CT arm and 135 in the B + CT arm) were included in this analysis. The median number of tumor assessments per patient was 4 (with a maximum of 16).

The final TK model adequately described the observed values. The model parameter estimates are reported in Table A1 in Appendix 1.

Although the typical values for  $k_{rg}$ ,  $k_s$ , and  $Y_0$  were common to all patients (CT and B + CT), the final model contained a  $\lambda$  parameter estimate specific to each type of chemotherapy (paclitaxel, PLD, or topotecan) in both CT and B + CT groups, respectively, indicating a differential rate to tumor regrowth and, thus, treatment effect duration, as suggested in Husain et al.<sup>22</sup>

The major steps of the model-building process are listed in Appendix 1. An illustration of the fit to individual data is presented in Appendix 1 (Figures A1 and A2) for a random sample of patients as well as the diagnostic plots supporting the final model assessment (Figures A3 and A4). The predictive performance of the TK model showed that the distribution of simulated and observed data at baseline and week 8 were closely matching (Figure A5, Appendix 1).

Individual tumor size metrics estimates are summarized in Table 2. Most of the patients experienced reduction in tumor size with a trend to a more profound shrinkage (mean *MaxSh* equal to -22% vs -5.8%) over a longer period of time (mean *TTG* equal to 17.5 vs 4.3 weeks) in CT + B compared with CT treated patients. The histogram of *ETS8* is provided in Figure A6, Appendix 1.

### Survival model

Out of the 275 patients retained to develop the TK model, 3 had missing *SLD* measures during the treatment period leading to retain only 272 patients in the OS analysis with 139 (51.1%) on the CT arm and 133 (48.9%) in the B + CT arm.

In the bootstrap-based covariate analysis of the Cox model (see Table A2 in Appendix 2, for more details), 2 sets of factors were found to be influential (Table 3) on time to death (or censoring): those describing the type and severity of disease (histology, presence of ascites, FIGO stage, ECOG stage) and those summarizing the tumor size dynamics ( $Y_0$ , *ETS8*). Negative *ETS8* values were associated with a prolongation of survival time compared with the reference (and vice versa).

Histology grade, line of therapy, baseline CA-125, time from diagnosis, age, and treatment group were not retained in the model after adjusting for the other effects.

The *c*-index plot (Figure 1) shows that ECOG score and serous vs non-serous subtypes have low contribution to the risk discrimination power on OS. Similar results were obtained with the integrated-Brier score showing sequential reduction from 0.162 for model 1 to 0.155, 0.145, and 0.139 for models 2, 3, and 4, 5, 6, while model 7 (titled "Treatment group" in Figure 1) was associated with a larger value (hence poorer performance) of 0.172.

**Table 3.** Parameter estimates ( $\theta$ ) of the final overall survival model.

	$exp(\theta)$	95%CI <sub><math>exp(\theta)</math></sub>
ECOG = 1, 2, 3 (ref. = 0)	1.01	[0.99, 1.02]
Serous = Yes (ref. = no)	0.77	[0.55, 1.06]
Presence of ascites (ref. = absence)	2.36	[1.71, 3.27]
FIGO stage		
I	1.28	[0.28, 5.76]
II	0.77	[0.14, 4.28]
III	1.72	[0.42, 7.11]
IV	2.98	[0.71, 12.6]
$log(Y_0)^a$ (mm)	1.38	[1.13, 1.68]
<i>ETS8</i> <sup>b</sup> (ref. = 0)	1.03	[1.02, 1.04]

Abbreviations: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; FIGO, Fédération Internationale de Gynécologie et d'Obstétrique; SLD, sum of longest diameters.

<sup>a</sup>  $Y_0$ : model-predicted SLD value at time of treatment onset.

<sup>b</sup> *ETS8*: model-predicted early tumor shrinkage (%) at week 8.

Using a reduced model involving only ascites, FIGO stage,  $Y_0$ , and *ETS8*, and fixing, for illustration,  $Y_0$  at 70 mm and FIGO stage at stage III, we show how various levels of reduction on tumor size at week 8 affect the survival time in patients with PROC in absence of ascites (Figure 2).

## Discussion

Tumor angiogenesis plays a pivotal role in the growth and metastasis of ovarian cancer because of the unique pattern of early dissemination of free-floating cells that form tumor implants in the peritoneal cavity. In this analysis, the time course of tumor size was studied for patients with PROC treated with chemotherapy with or without bevacizumab, a potent anti-angiogenesis agent.

The TK modeling demonstrated that bevacizumab brought a deeper anti-tumor response, with a median value for the (absolute) maximum shrinkage of 22% in the combination arm compared with 6% in the CT arm, and also prolonged the shrinkage as illustrated by a time to growth extending from 4 weeks in the CT arm to more than 17 weeks in the combination arm.

However, the observed benefits of B on TK did not translate into significant OS benefit (HR = 0.85 [0.66, 1.08] in favor of B + CT<sup>2</sup>).

Several reasons can explain the disconnection between TK and OS. OS represents death events from any causes and not only due to disease progression; the variability observed in OS could not be explained solely by TK. In addition, the primary endpoint of the AURELIA study was PFS and the study was not powered to detect differences in OS. A recent meta-analysis

of 5 large phase III trials in recurrent OC actually concluded at a benefit of B + CT compared with CT alone.<sup>23</sup>

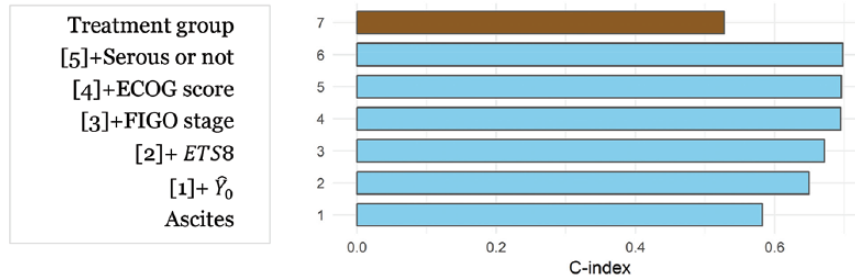
An important covariate influencing OS was the presence or absence of ascites at baseline, a well-known prognostic factor. This finding echoes the results of the analysis run by the NRG Oncology/GOG study<sup>24</sup> and emphasizes the complex cause of ovarian cancers. Although other baseline covariates were considered to be tested in the Cox regression, CA-125 was not. Indeed, CA-125 has been considered as a potential surrogate endpoint across all subpopulation of OC. However, in a recent study, neither disease progression by CA-125 nor baseline CA-125 was found to be predictive of disease progression in patients with PROC showing the limitations of CA-125.<sup>25</sup>

To relate TK to survival, we assumed that any drug effect on tumor burden would translate, or at least correlate with, a change in the risk of death. Because these 2 endpoints are correlated, including post treatment TK in a time-to-event model would have led to biased results. To circumvent this issue, we used a landmark approach,<sup>26</sup> and evaluated the correlation between early tumor shrinkage at week 8 and OS among patients who had survived at least 8 weeks after treatment initiation. Choosing a landmark is also associated with a loss of power due to the reduced sample size as compared with analyzing the full analysis set.<sup>27</sup> However, the risk for erroneous conclusions was considered low in our case due to the relatively large number of subjects with data available for our analysis. An alternative approach would have been to develop a joint model of tumor size dynamics and OS. Although joint models involving a longitudinal linear sub-model to describe the biomarker time dynamics have been applied multiple times in oncology, the methodology to deal with longitudinal non-linear sub-models (like the one describing TK) is in its infancy.<sup>28</sup>

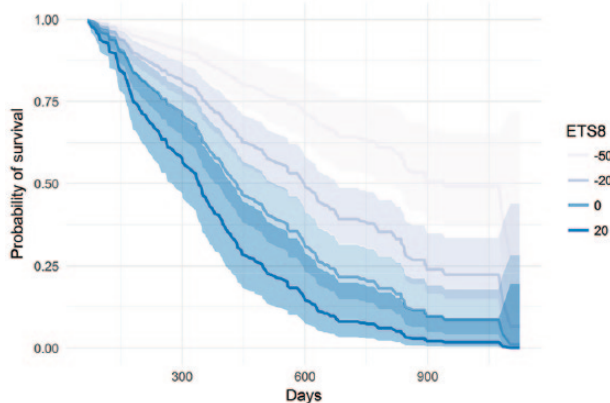
Overall survival remains the most clinically meaningful endpoint in oncology and the US Food and Drug Administration (FDA) gold standard for drug approval. The identification of early markers or assessments that can inform the likelihood of an OS advantage being achieved is crucial. As the opposite of baseline covariates (FIGO stage, presence of ascites, ECOG), *ETS8* offers the advantage to be an early predictive marker that not only correlates to OS but also informs on the potential benefits of the studied drug.

A similar but more comprehensive model-based approach was recently proposed by Zecchin et al<sup>16</sup> to assess the effect of carboplatin monotherapy or in combination with gemcitabine in patients with PROC. In their final OS model, both target lesion size and appearance of new lesions were significant covariates in addition to the observed baseline SLD and ECOG status at enrolment. The AURELIA trial was evaluated using RECIST1.0 that did not account for the appearance of new lesions. Therefore, new lesion could not be considered for inclusion in our survival analysis.

In conclusion, this is the first analysis establishing that early tumor shrinkage is predictive of OS in patients with PROC. This modeling framework could effectively help to simulate and optimize future trials in PROC population.



**Figure 1.** Concordance index (overall c-index) obtained with different OS models.



**Figure 2.** Predicted probability of survival for various levels of *ETS8* (%) in absence of ascites. Shaded areas correspond to the parametric 95% confidence interval around the mean probability.

## Acknowledgements

The authors thank participants and investigators of the AURELIA study.

## Author Contributions

Both authors contributed to the conception of the work, participated in the analysis and interpretation of data, performed a critical review of its content, and approved the final manuscript.

## ORCID iDs

Alexandre Sostelly  <https://orcid.org/0000-0002-7300-979X>

François Mercier  <https://orcid.org/0000-0002-5685-1408>

## REFERENCES

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61:69–90.
- Pujade-Lauraine E, Hilpert F, Weber B, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: the AURELIA open-label randomized phase III trial. *J Clin Oncol*. 2014;32:1302–1308.
- Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol*. 2009;27:2874–2880.
- Villaruz LC, Socinski MA. The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. *Clin Cancer Res*. 2013;19:2629–2636.
- Sridhara R, Mandrekar SJ, Dodd LE. Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res*. 2013;19:2613–2620.
- Campigotto F, Weller E. Impact of informative censoring on the Kaplan-Meier estimate of progression-free survival in phase II clinical trials. *J Clin Oncol*. 2014;32:3068–3074.
- Zeng L, Cook RJ, Wen L, Boruvka A. Bias in progression-free survival analysis due to intermittent assessment of progression. *Stat Med*. 2015;34:3181–3193.
- Tate Thigpen J. Contemporary phase III clinical trial endpoints in advanced ovarian cancer: assessing the pros and cons of objective response rate, progression-free survival, and overall survival. *Gynecol Oncol*. 2015;136:121–129.
- Blumenthal GM, Pazdur R. Response rate as an approval end point in oncology: back to the future. *JAMA Oncol*. 2016;2:780–781.
- Claret L, Girard P, Hoff PM, 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;27:4103–4108.
- Hansson EK, Ma G, Amantea MA, et al. PKPD modeling of predictors for adverse effects and overall survival in sunitinib-treated patients with GIST. *CPT Pharmacometrics Syst Pharmacol*. 2013;2:e85.
- Maitland ML, Wu K, Sharma MR, et al. Estimation of renal cell carcinoma treatment effects from disease progression modeling. *Clin Pharmacol Ther*. 2013;93:345–351.
- Claret L, Mercier F, Houk BE, Milligan PA, Bruno R. Modeling and simulations relating overall survival to tumor growth inhibition in renal cell carcinoma patients. *Cancer Chemother Pharmacol*. 2015;76:567–573.
- Wang Y, Sung C, Dartois C, 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;86:167–174.
- Bruno R, Mercier F, Claret L. Evaluation of tumor size response metrics to predict survival in oncology clinical trials. *Clin Pharmacol Ther*. 2014;95:386–393.
- Zecchin C, Gueorguieva I, Enas NH, Friberg LE. Models for change in tumour size, appearance of new lesions and survival probability in patients with advanced epithelial ovarian cancer. *Br J Clin Pharmacol*. 2016;82:717–727.
- Claret L, Gupta M, Han K, et al. Evaluation of tumor-size response metrics to predict overall survival in Western and Chinese patients with first-line metastatic colorectal cancer. *J Clin Oncol*. 2013;31:2110–2114.
- Venables WN, Ripley BD. *Modern Applied Statistics With S*. 4th ed. New York, NY: Springer; 2002.
- Austin PC, Tu JC. Bootstrap methods for developing predictive models. *Am Stat*. 2004;58:131–137.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
- Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *J Stat Softw*. 2012;50:1–23.
- Husain A, Wang Y, Hanker LC, et al. Independent radiologic review of AURELIA, a phase 3 trial of bevacizumab plus chemotherapy for platinum-resistant recurrent ovarian cancer. *Gynecol Oncol*. 2016;142:465–470.
- Wu YS, Shui L, Shen D, Chen X. Bevacizumab combined with chemotherapy for ovarian cancer: an updated systematic review and meta-analysis of randomized controlled trials. *Oncotarget*. 2017;8:10703–10713.
- Ferriss JS, Java JJ, Bookman MA, et al. Ascites predicts treatment benefit of bevacizumab in front-line therapy of advanced epithelial ovarian, fallopian tube and peritoneal cancers: an NRG Oncology/GOG study. *Gynecol Oncol*. 2015;139:17–22.
- Lindemann K, Kristensen G, Mirza MR, et al. Poor concordance between CA-125 and RECIST at the time of disease progression in patients with platinum-resistant ovarian cancer: analysis of the AURELIA trial. *Ann Oncol*. 2016;27:1505–1510.
- Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1:710–719.
- Dafni U. Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes*. 2011;4:363–371.
- Desmees S, Mentre F, Veyrat-Follet C, Guedj J. Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: a comparison by simulation of two-stage and joint approaches. *AAPS J*. 2015;17:691–699.
- Stein WD, Gulley JL, Schlom J, 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;17:907–917.
- Claret L, Gupta M, Han K, et al. Evaluation of tumor-size response metrics to predict overall survival in western and Chinese patients with first-line metastatic colorectal cancer. *J Clin Oncol*. 2013;31:2110–2114.

## Appendix 1

Additional details on tumor kinetic models development.

### Model building steps

MODEL NUMBER	DESCRIPTION	AIC	SUCCESSFUL ESTIMATION OF STANDARD ERRORS
Model 1	Stein Model <sup>29</sup>	8551	Y
Model 2	Wang Model <sup>14</sup>	8935	Y
Model 3	Claret Model <sup>30</sup>	8545	Y
Model 4	Model 3 + Arm as binary covariate (B + CT vs CT) on $k_s$	8529	Y
Model 5	Model 3 + Arm as binary covariate (B + CT vs CT) on $\lambda$	8534	Y
Model 6	Model 3 + Arm as binary covariate (B + CT vs CT) on $\lambda$ and $k_s$	8536	Y
Model 7	Model 3 + Each treatment as binary covariate (B + paclitaxel, B + PLD, B + topotecan, paclitaxel, PLD, topotecan) on $k_s$	8517	Y
Model 8	<b>Model 3 + Each treatment as binary covariate (B + paclitaxel, B + PLD, B + topotecan, paclitaxel, PLD, topotecan) on <math>\lambda</math></b>	<b>8513</b>	<b>Y</b>
Model 9	Model 3 + Each treatment as binary covariate (B + paclitaxel, B + PLD, B + topotecan, paclitaxel, PLD, topotecan) on $\lambda$ and $k_s$	8487	N

Model 9 was not retained because of the unsuccessful estimation of standard errors reflecting model over-parametrization. Abbreviations: AIC, Akaike information criteria; B, bevacizumab; CT, chemotherapy; PLD, pegylated liposomal doxorubicin.

**Table A1.** TK model parameter estimates.

PARAMETER	GROUP	ESTIMATE (RSE <sup>A</sup> %)	%IIV <sup>B</sup> (RSE <sup>A</sup> %)	SHRINK. <sup>C</sup> (%)	MEDIAN BOOTSTRAP VALUES <sup>D</sup> [90% CONFIDENCE INTERVAL]
SLD increase rate, day <sup>-1</sup> ( $k_{rg}$ )		$4.98 \times 10^{-4}$ (25)	138.5 (10)	38.4	$4.8 \times 10^{-4}$ [ $2.6 \times 10^{-4}$ , $8 \times 10^{-4}$ ]
SLD decrease rate, day <sup>-1</sup> ( $k_s$ )		$6.98 \times 10^{-3}$ (18)	85.6 (11)	41.8	$7.12 \times 10^{-3}$ [ $5.1 \times 10^{-3}$ , $1.0 \times 10^{-2}$ ]
Baseline, mm ( $\hat{Y}_0$ )		55.5 (5)	82.8 (4)	6.0	55.4 [50.68, 60.23]
Resistance rate, day <sup>-1</sup> ( $\lambda$ )	Paclitaxel + B	$3.87 \times 10^{-2}$ (45)	149.6 (13)	42.9	$4.6 \times 10^{-2}$ [ $2.2 \times 10^{-2}$ , $2.23 \times 10^{-1}$ ]
	PLD + B	$7.25 \times 10^{-3}$ (32)			$7.62 \times 10^{-3}$ [ $3.81 \times 10^{-3}$ , $1.31 \times 10^{-2}$ ]
	Topotecan + B	$3.28 \times 10^{-2}$ (38)			$3.74 \times 10^{-2}$ [ $1.7 \times 10^{-2}$ , $9.5 \times 10^{-2}$ ]
	Paclitaxel	$7.99 \times 10^{-2}$ (40)			$9.06 \times 10^{-2}$ [ $3.5 \times 10^{-2}$ , $2.5 \times 10^{-1}$ ]
	PLD	$2.48 \times 10^{-2}$ (32)			$2.56 \times 10^{-2}$ [ $1.4 \times 10^{-2}$ , $4.4 \times 10^{-2}$ ]
	Topotecan	$1.30 \times 10^{-1}$ (59)			$1.62 \times 10^{-1}$ [ $6.0 \times 10^{-2}$ , $8.0 \times 10^{-1}$ ]
Residual error, mm		8.41 (8)	NA	26.3	8.42 [7.11, 9.78]

Model parameter estimates and their respective 90% confidence intervals generated by bootstrapping are presented in Table A1. Median bootstrap values were close to the values estimated from the original dataset indicating a small bias. Parameters were estimated with adequate precision and low shrinkage.

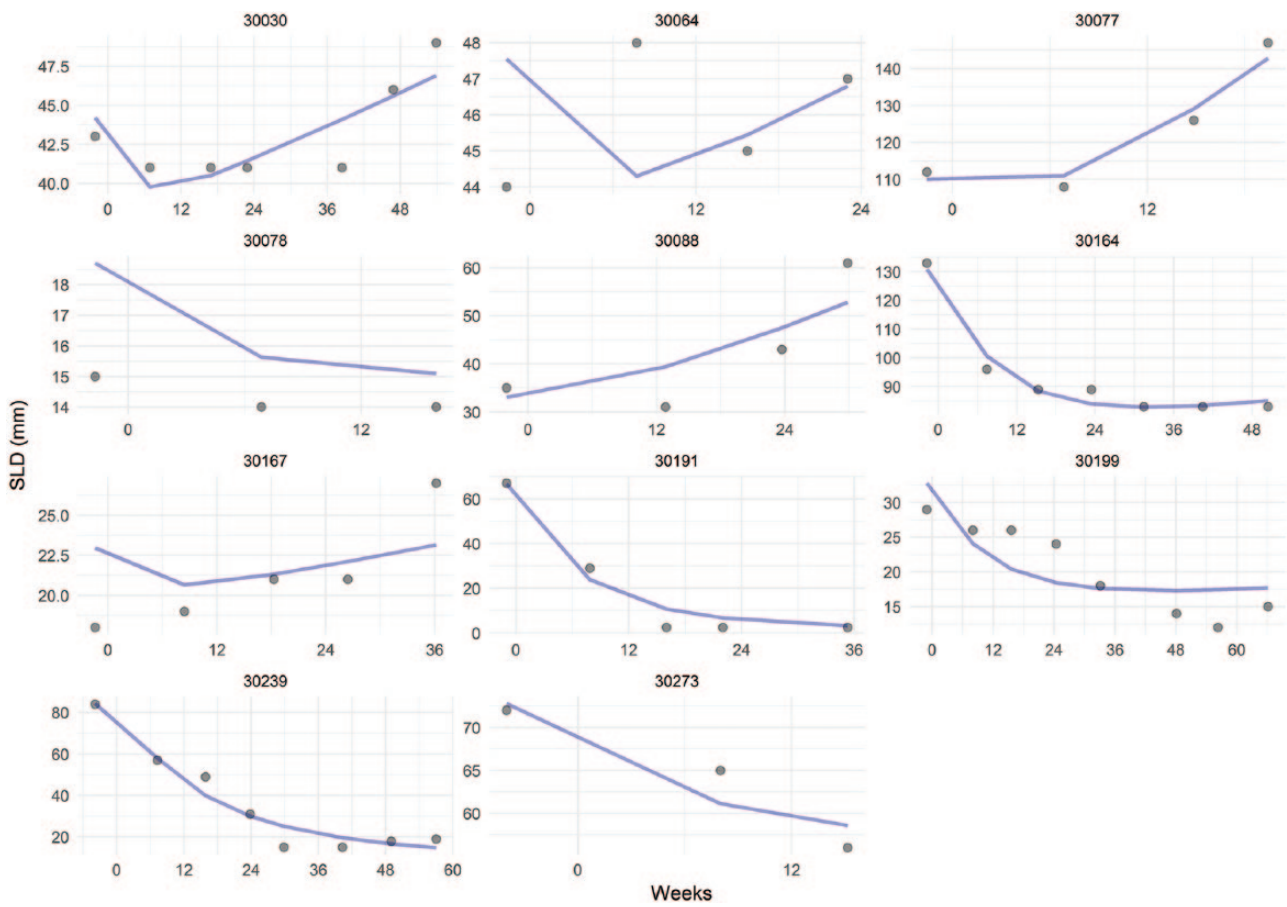
<sup>A</sup>RSE: relative standard error.

<sup>B</sup>%IIV: inter-individual variability expressed as % coefficient of variation.

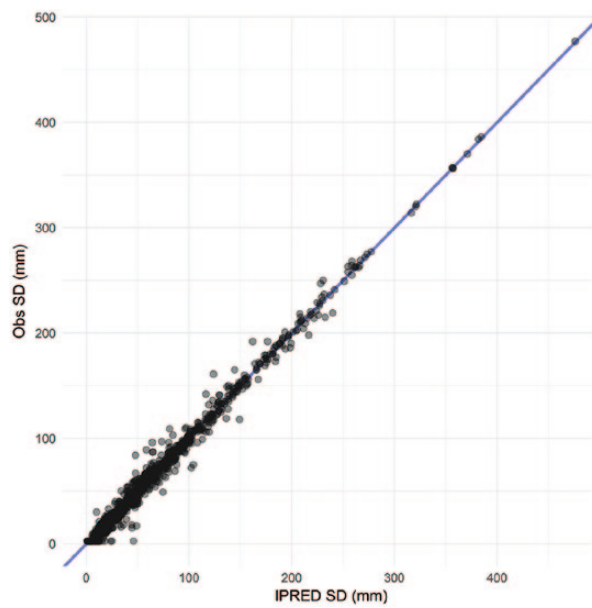
<sup>C</sup>Shrink.: parameter estimate shrinkage.

<sup>D</sup>10% trimmed median derived from 100 bootstrap replicates.

GoF of final TK model



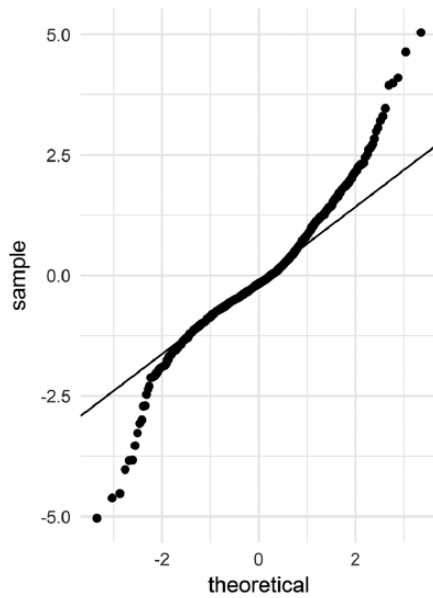
**Figure A1.** Goodness-of-fit of the final model to the SLD values for a random sample of 9 patients from the AURELIA study. The curve depicts the model predicted SLD over time. The dots are the observations. SLD indicates sum of longest diameters.



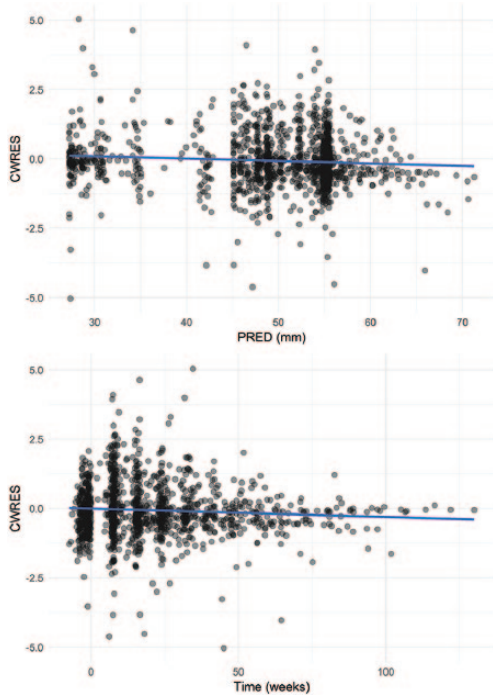
**Figure A2.** Goodness-of-fit of the final model to the SD values: observations vs individual model predicted SLD. SLD indicates sum of longest diameters.

The assumption of normality of the observations made using the maximum likelihood approach implemented in the FOCE method was tested by inspecting the quantile-quantile plots of the conditional weighted residuals (CWRES) against theoretical standard normal quantiles (Figures A3 and A4).

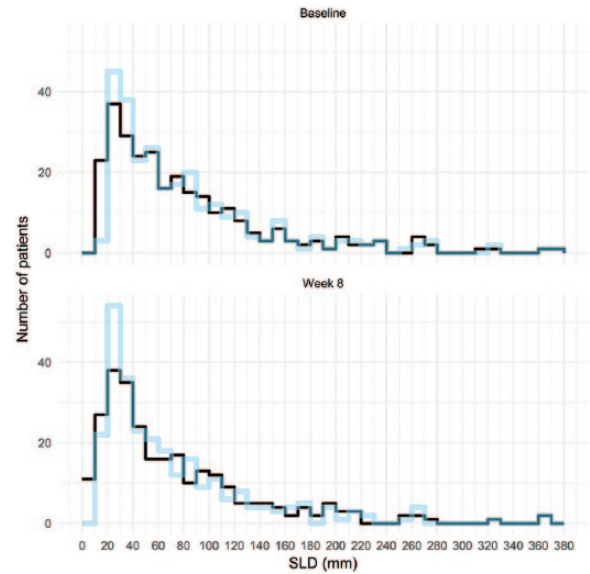
Figure A3 confirms that the assumption of normality of the observations made during estimation is verified. Figure A4 shows no model misspecification.



**Figure A3.** Goodness-of-fit of the final model to the SD values: Q-Q plot of CWRES.

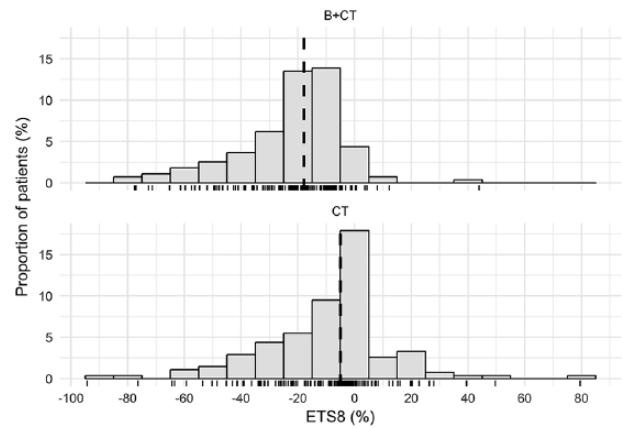


**Figure A4.** Goodness-of-fit of the final model to the SD values: CWRES.



**Figure A5.** Distributions of observed vs predicted SLD values obtained from the “baseline” scan and from the first scan on treatment (at week 8).

The predictive performance of the TK model (evaluated using the visual predictive check) was considered adequate as the distribution of simulated and observed data at baseline and week 8 were closely matching (Figure A5).



**Figure A6.** Histogram of ETS8.



## Appendix 2

**Table A2.1.** Bootstrap-based covariate analysis.

```

> cox<-coxph(Surv(OS,
OSC==0)~as.factor(TRTC)+as.factor(BECOG)+as.factor(HGRAD)+
  as.factor(SEROUS)+as.factor(LTHER)+as.factor(ASCITES)+
  as.factor(FIGO1)+as.factor(FIGO2)+as.factor(FIGO3)+as.factor(FIGO4)+
  cAGE+as.factor(cDIAG)+BLCA125+
  logPREDBASE+ETS8, landf2)

> set.seed(1447)
> library(bootStepAIC, lib.loc="C:\\MyR\\MyLibs")
> bootStepAIC::boot.stepAIC(cox, df, direction="both", alpha=0.001)
> detach("package:bootStepAIC", unload=TRUE)

> bootStepAIC::boot.stepAIC(cox, df, direction="both", alpha=0.001)

the model fit failed in 15 bootstrap samples

Summary of Bootstrapping the 'stepAIC()' procedure for

Call:
coxph(formula = Surv(OS, OSC == 0) ~ as.factor(TRTC) + as.factor(BECOG) +
  as.factor(HGRAD) + as.factor(SEROUS) + as.factor(LTHER) +
  as.factor(ASCITES) + as.factor(FIGO1) + as.factor(FIGO2) +
  as.factor(FIGO3) + as.factor(FIGO4) + cAGE + as.factor(cDIAG) +
  BLCA125 + logPREDBASE + ETS8, data = landf2)

Bootstrap samples: 85
Direction: both
Penalty: 2 * df

Covariates selected
                                (%)

as.factor(ASCITES)                100.00
ETS8                               100.00
logPREDBASE                       100.00
as.factor(BECOG)                   92.94
as.factor(FIGO4)                   72.94
as.factor(FIGO2)                   55.29
as.factor(FIGO3)                   54.12
as.factor(SEROUS)                  50.59
as.factor(FIGO1)                   49.41
as.factor(TRTC)                    30.59
as.factor(HGRAD)                   29.41
as.factor(LTHER)                   28.24
as.factor(cDIAG)                   24.71
cAGE                               18.82
BLCA125                            8.24

Coefficients Sign
                                + (%)          - (%)

as.factor(ASCITES)1                100.00          0.00
as.factor(BECOG)1                  100.00          0.00
as.factor(BECOG)99                  100.00          0.00
as.factor(FIGO4)1                   100.00          0.00
ETS8                                100.00          0.00
logPREDBASE                         98.82           1.18
as.factor(LTHER)2                   95.83           4.17
as.factor(HGRAD)1                   84.00          16.00
cAGE                                 81.25          18.75
as.factor(FIGO3)1                   50.00          50.00
as.factor(FIGO1)1                   45.24          54.76
BLCA125                              28.57          71.43
as.factor(TRTC)CT+BV                15.38          84.62
as.factor(cDIAG)1                   9.52           90.48
as.factor(FIGO2)1                   8.51           91.49
as.factor(SEROUS)1                  0.00          100.00

```

(Continued)

Table A2.1. (Continued)

Stat Significance

	(%)
as.factor(ASCITES)1	96.47
ETS8	83.53
as.factor(BECOG)1	60.76
as.factor(FIGO4)1	51.61
as.factor(FIGO3)1	32.61
logPREDBASE	31.76
as.factor(SEROUS)1	13.95
as.factor(FIGO2)1	12.77
as.factor(HGRAD)1	8.00
as.factor(cDIAG)1	4.76
as.factor(FIGO1)1	4.76
as.factor(LTHER)2	4.17
as.factor(TRTC)CT+BV	3.85
as.factor(BECOG)99	1.54
BLCA125	0.00
cAGE	0.00

The stepAIC() for the original data-set gave

Call:

```
coxph(formula = Surv(OS, OSC == 0) ~ as.factor(BECOG) + as.factor(SEROUS) +
  as.factor(ASCITES) + as.factor(FIGO2) + as.factor(FIGO4) +
  logPREDBASE + ETS8, data = landf2)
```

	coef	exp(coef)	se(coef)	z	p
as.factor(BECOG)1	0.51782	1.67836	0.15223	3.40	0.00067
as.factor(SEROUS)1	-0.23656	0.78934	0.16049	-1.47	0.14050
as.factor(ASCITES)1	0.79617	2.21703	0.16469	4.83	1.3e-06
as.factor(FIGO2)1	-0.80568	0.44679	0.51075	-1.58	0.11469
as.factor(FIGO4)1	0.55778	1.74680	0.16519	3.38	0.00073
logPREDBASE	0.26828	1.30772	0.09706	2.76	0.00571
ETS8	0.02279	1.02305	0.00447	5.10	3.3e-07

Likelihood ratio test=105 on 7 df, p=0  
 n= 270, number of events= 195  
 (77 observations deleted due to missingness)

Stepwise Model Path  
 Analysis of Deviance Table

Initial Model:

```
Surv(OS, OSC == 0) ~ as.factor(TRTC) + as.factor(BECOG) + as.factor(HGRAD) +
  as.factor(SEROUS) + as.factor(LTHER) + as.factor(ASCITES) +
  as.factor(FIGO1) + as.factor(FIGO2) + as.factor(FIGO3) +
  as.factor(FIGO4) + cAGE + as.factor(cDIAG) + BLCA125 + logPREDBASE +
  ETS8
```

Final Model:

```
Surv(OS, OSC == 0) ~ as.factor(BECOG) + as.factor(SEROUS) + as.factor(ASCITES) +
  as.factor(FIGO2) + as.factor(FIGO4) + logPREDBASE + ETS8
```

Step	Df	Deviance Resid.	Df	Resid. Dev	AIC
1	1		255	1813.633	1843.633
2 - BLCA125	1	0.02559867	256	1813.658	1841.658
3 - as.	1	0.02846304	257	1813.687	1839.687
factor(FIGO1)	1	0.11863696	258	1813.805	1837.805
4 - as.factor(TRTC)	1	0.19574818	259	1814.001	1836.001
5 - cAGE	1	0.25908138	260	1814.260	1834.260
6 - as.	1	0.50917455	261	1814.769	1832.769
factor(HGRAD)	1	0.81839384	262	1815.588	1831.588
7 - as.		0.35617270	263	1815.944	1829.944
factor(FIGO3)					
8 - as.					
factor(cDIAG)					
9 - as.					
factor(LTHER)					

There were 50 or more warnings (use warnings[] to see the first 50)