

## Tumor Burden Modeling Versus Progression-Free Survival for Phase II Decision Making

Lee D. Kaiser

### Abstract

Randomized Phase II oncology trial endpoints for decision making include both progression-free survival (PFS) and change in tumor burden as measured by the sum of longest diameters (SLD) of the target lesions. In addition to observed SLD changes, tumor shrinkage and growth parameters can be estimated from the patient-specific SLD profile over time. The ability of these SLD analyses to identify an active drug is contrasted with that of a PFS analysis through the simulation of Phase II trials via resampling from each of 6 large, Phase II and III trials, 5 of which were positive and one negative. From each simulated Phase II trial, a *P* value was obtained from 4 analyses—a log-rank test on PFS, a Wilcoxon rank-sum test on the minimum observed percentage change from baseline in SLD, and 2 nonlinear, mixed-effects model analyses of the SLD profiles. All 4 analyses led to approximately uniformly distributed *P* values in the negative trial. The PFS analysis was the best or nearly the best analysis in the other 5 trials. In only one of the positive studies did the modeling analysis outperform the analysis of the minimum SLD. In conclusion, for the decision to start a Phase III trial based on the results of a randomized Phase II trial of an oncology drug, PFS appears to be a better endpoint than does SLD, whether analyzed through simple SLD endpoints, such as the minimum percentage change from baseline, or through the modeling of the SLD time course to estimate tumor dynamics. *Clin Cancer Res*; 19(2); 314–9. ©2012 AACR.

### Introduction

There are many possible designs of a Phase II oncology trial (1), and randomized Phase II trials are often recommended (2–5). In such randomized trials, patients' tumors are assessed through radiographic imaging, and Response Evaluation Criteria in Solid Tumors (RECIST) criteria (6) applied for determination of tumor burden, tumor response, and progression-free survival (PFS), defined as the earlier of the occurrence of progressive disease or death. Even though overall survival is the gold standard for demonstration of efficacy in Phase III, it is infrequently the primary endpoint in Phase II owing to large sample size and long follow-up requirements (7). Rather, a decision to continue to Phase III with the experimental therapy typically relies on the comparison of response rates and PFS between the randomized treatment arms (7, 8).

There have been recent suggestions that the percentage change from baseline in tumor burden, as measured by the sum of longest diameters (SLD) of the target lesions, can be used for the assessment of comparative efficacy in randomized trials (9–18). The proposals fall into 2 categories. One (9, 10) suggests use of the percentage change from

baseline in SLD, denoted as SLD%, as a better endpoint than response rate, which is essentially a dichotomized minimum SLD%. The other category (11–18) models SLD as a function of time. Buyse and colleagues (11) suggest that "a model that uses all tumor size measurements for each patient may be preferable to a model that uses PFS, given that the latter design makes less efficient use of these data." Modeling of SLD has also been used to

- predict overall survival as a function of early SLD% and other variables, which is then used to predict Phase III outcomes (12–14),
- assess the response of tumors to time-varying dose levels of an experimental drug (15), and
- evaluate the relative efficacy of study treatments through the comparison of fitted tumor growth parameters between treatment arms (16–18).

Tumor burden analyses have been compared with PFS analyses in their ability to lead to the correct decision about the initiation of a Phase III trial. Through the simulation of Phase II studies from 6 large completed trials, it was shown (19) that the analysis of PFS in a randomized Phase II trial generally leads to better decisions about starting a Phase III trial than does the comparison between treatment groups of simple, per-patient tumor burden endpoints, such as the minimum or the last SLD%. However, in the simulation of Phase II studies from 1 positive Phase III trial (20), an endpoint equivalent to SLD% at the first postbaseline assessment performed better than PFS in leading to the

**Author's Affiliation:** Genentech, South San Francisco, California

**Corresponding Author:** Lee Kaiser, 1 DNA Way, MS 452a, South San Francisco, CA 94080. Phone: 650-225-5925; Fax: 650-225-5065; E-mail: lkaiser@gene.com

**doi:** 10.1158/1078-0432.CCR-12-2161

©2012 American Association for Cancer Research.

**Table 1.** Study description

Study <sup>a</sup>	Indication	Treatments	Tumor assessment frequency	Response rates (%) control, experimental	PFS HR
AVF2107g (21) <i>n</i> = 813/750	First line CRC	IFL +/- bevacizumab	Every 6 weeks for 24 weeks, then every 12 weeks	34.8, 44.8; <i>P</i> = 0.004	0.54; <i>P</i> < 0.0001
AVF2119g (22) <i>n</i> = 462/412	Second line BC	Capecitabine +/- bevacizumab	Every 6 weeks for 24 weeks, then every 9 weeks	9.1, 19.8; <i>P</i> = 0.001	0.98; <i>P</i> = 0.86
AVF2192g (23) <i>n</i> = 209/189	First line CRC	5-FU/LV +/- bevacizumab	Every 8 weeks	15.2, 26.0; <i>P</i> = 0.055	0.50; <i>P</i> = 0.0002
SO14999 (24) <i>n</i> = 511/477	BC, 33% first line, otherwise later lines	Docetaxel +/- capecitabine	Every 6 weeks for 48 weeks, then every 12 weeks	30, 42; <i>P</i> = 0.006	0.65 <sup>b</sup> ; <i>P</i> = 0.0001
BR21 (25) <i>n</i> = 731/525	Second line NSCLC	Erlotinib vs. placebo	Every 8 weeks	<1, 8.9; <i>P</i> < 0.001	0.61; <i>P</i> < 0.001
H0648g (26) <i>n</i> = 469/361	First line HER2+ BC	AC +/- trastuzumab pac +/- trastuzumab	An 8-week assessment, then every 12 weeks	32, 50; <i>P</i> < 0.001	0.51 <sup>b</sup> ; <i>P</i> < 0.001

Abbreviations: CRC, colorectal cancer; BC, breast cancer; IFL, irinotecan, 5-FU, leucovorin; LV, leucovorin; AC, anthracycline, cyclophosphamide; pac, paclitaxel.

<sup>a</sup>*n* = the number of patients randomized in the study over the number of patients remaining for the analysis after data processing. For study AVF2107g, *n* = 813 excludes 110 patients randomized to a 5-FU/LV + bevacizumab arm that was dropped from the study. Other patient exclusion reasons are described in the text.

<sup>b</sup>HR for analysis of time to progression.

correct decision to continue to Phase III, but at the cost of a larger false-positive rate for a negative trial.

Modeling of the SLD time course in a Phase II trial might lead to better decisions than the analysis of these simple SLD endpoints, with perhaps even a consistent advantage over PFS. The present manuscript evaluates this possibility.

## Materials and Methods

Phase II trials were simulated from the 6 completed trials listed in Table 1 (19, 21–26). Tumors were assessed at intervals of 6, 8, 9, or 12 weeks, depending on the study. Progressive disease was determined using RECIST criteria for the bevacizumab and erlotinib studies and using World Health Organization (WHO) criteria (27) for the capecitabine trial and modified WHO for the trastuzumab trial. The sum of the longest diameters across the target lesions was determined for the bevacizumab and erlotinib studies and across the marker lesions for the trastuzumab and capecitabine studies. All studies but AVF2119g were positive for PFS, and all studies were positive or nearly so for overall response rate. All studies but AVF2119g and AVF2192g were positive for overall survival. Study AVF2119g did not lead to the successful registration of the drug, and is a negative study for the present analysis.

Patients were included in the analyses if they had a baseline tumor assessment, received at least 1 dose of study medication, and had at least 1 postbaseline tumor assessment. The effect of excluding patients with no postbaseline tumor assessment is addressed in the discussion. While there is large variability in the size, enrollment duration,

and patient follow-up in published randomized Phase II trials, the simulated Phase II trials here had intermediate characteristics of 100 patients enrolled uniformly over 1 year, with 50 of these patients selected at random from the control arm of the parent study and 50 from the experimental arm. Each simulated trial included all SLD data available on the selected patients through 6 months after the last patient's enrollment time. If a patient's PFS in a simulated trial was after this 6-month cutoff, then the value was censored at the time of this cutoff.

The comparison of PFS between treatment groups within each replicate was via a log-rank test. The minimum SLD% was compared between treatments with a Wilcoxon rank-sum test. The SLD modeling comparisons followed from a nonlinear, mixed-effects model of SLD in cm (12, 16):

$$SLD_{ijk} = A_{ij} \{ \exp(-S_{ij} t_{ijk}) + \exp(G_{ij} t_{ijk}) - 1 \} + e_{ijk}, [1]$$

where

*i* = treatment group, either 0 for control, or 1 for experimental,

*j* = patient within treatment group,

*k* = observation within patient *j* and treatment group *i*,

*A*<sub>*ij*</sub> = baseline SLD (cm) for patient *ij*,

*t*<sub>*ijk*</sub> = time of the *k*<sup>th</sup> tumor assessment for patient *ij*, with *t*<sub>*ij1*</sub> = 0

*S*<sub>*ij*</sub> = a shrinkage parameter (1/time) for patient *ij*,

*G*<sub>*ij*</sub> = a growth parameter (1/time) for patient *ij*,

*e*<sub>*ijk*</sub> = a random error (cm).

Distributional assumptions are as follows: log(*A*<sub>*ij*</sub>), log(*S*<sub>*ij*</sub>), and log(*G*<sub>*ij*</sub>) are independent between patients and

distributed as multivariate normal with means  $\log(\alpha)$ ,  $\log(\xi_i)$ , and  $\log(\gamma_i)$ , respectively, and with an arbitrary variance-covariance matrix, except that the covariance between  $\log(A_{ij})$  and each of  $\log(S_{ij})$  and  $\log(G_{ij})$  is assumed to be 0. This 0 covariance setting is consistent with the findings that SLD% is roughly independent of baseline SLD (19) and that the estimated shrinkage and growth parameters are independent of baseline SLD (18). The  $e_{ijk}$  are independent and identically distributed as normal, with 0 mean and common variance. Parameter estimation was conducted with Monolix 3.2 (28, 29).

One test for a treatment effect in this model is a Wald test, with 2 degrees of freedom, of  $H_0: \xi_0 = \xi_1$  and  $\gamma_0 = \gamma_1$ . A treatment that resulted in generally smaller SLD values after baseline should be better detected with a test with 1 degree of freedom, and for this the following test was used. The shrinkage terms were constrained to be equal between treatment arms, and the difference between arms in growth

parameters was tested with a Wald test. Other SLD-based treatment arm comparisons conducted but not presented are described in the discussion.

Because nonlinear, mixed-effects modeling is time consuming per simulated trial, the assessment was limited to 100 simulated Phase II trials per parent study. For the visual assessment of the treatment comparisons via the SLD analyses and the PFS analysis, the empirical cumulative distribution function (CDF) of the 100 2-sided  $P$  values was plotted for each analysis method and each parent clinical trial. The empirical distribution function provides the percentage of  $P$  value results less than or equal to any given value.

## Results

Figure 1 plots the observed SLD versus the model-predicted SLD for 1 Phase II replicate per parent trial. The fit of the model to the data looks satisfactory, and in particular, the choice of an additive error term is supported.

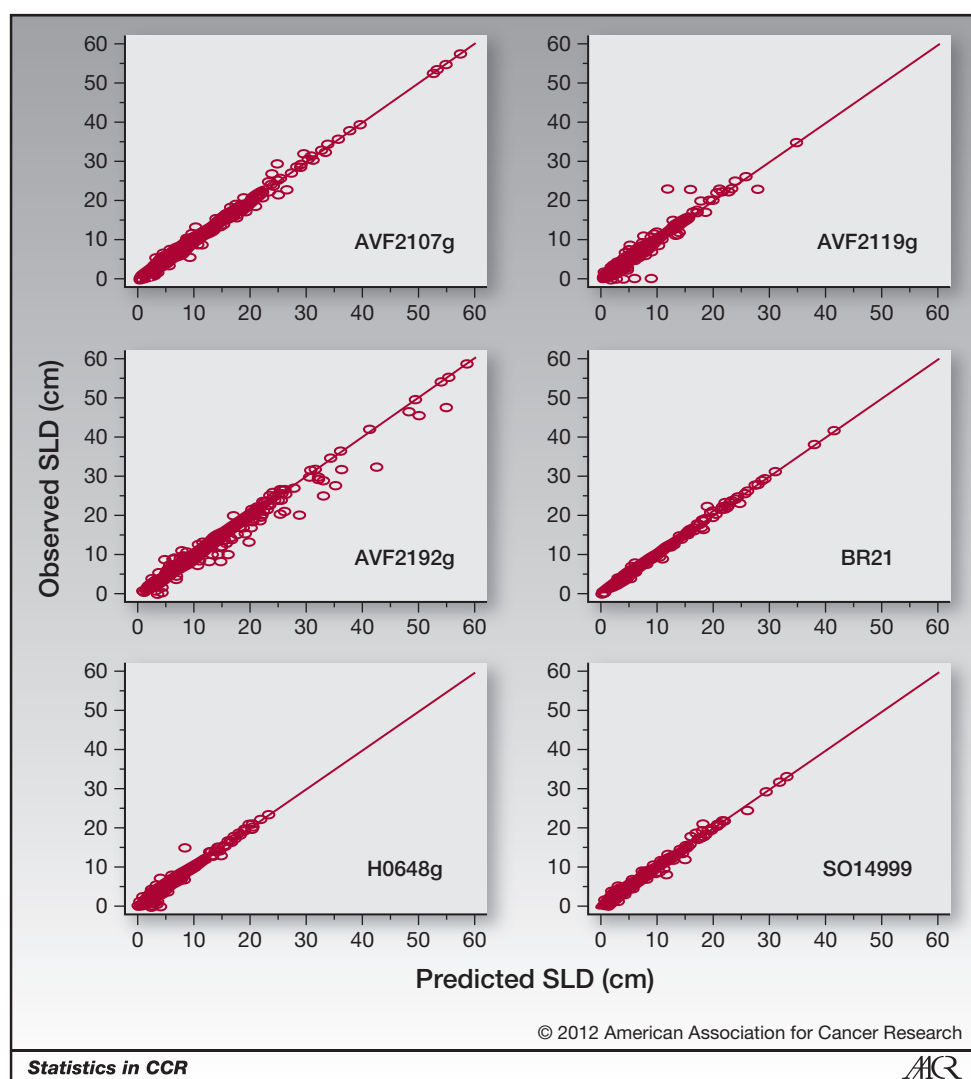


Figure 1. Observed versus predicted SLD from the first Phase II replicate for each parent trial.

Figure 2. Empirical distribution functions of  $P$  values across the 100 simulated Phase II studies for each test for a treatment difference and for each parent trial.

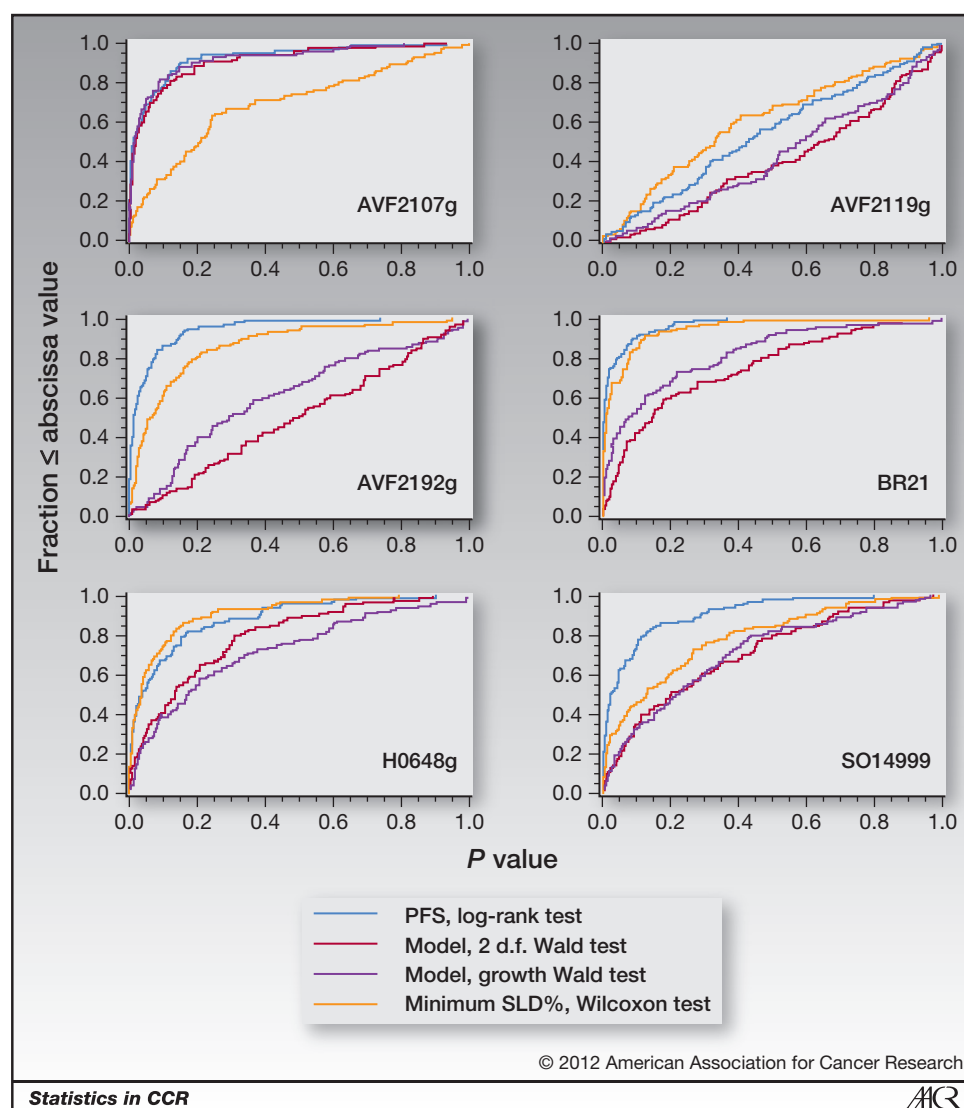


Figure 2 contains the empirical CDFs for the 100  $P$  values from each test for each of the parent studies. For a positive study, a better test for treatment effect has an empirical CDF that rises quickly from the origin, indicating many replicates with small  $P$  values, and a poorer test has an empirical CDF closer to the  $45^\circ$  line, which corresponds to a uniform distribution. As an example of reading these curves, approximately 30% of the  $P$  values from the minimum SLD% analysis in study AVF2107g were less than or equal to 0.10. Also, approximately 75% of the  $P$  values from the PFS analysis in study SO14999 were less than or equal to 0.10. Depending on the  $P$  value cutoff chosen for determining a positive study, these graphs can be used to assess the specificity of the analysis method for study AVF2119g and the sensitivity of the method for the other studies.

In the negative AVF2119g study the distribution of the  $P$  values from each test is approximately uniform, so that none of the analyses would be expected to lead to greatly increased Type I error rates. However, there was a significant

effect of treatment on the response rate in the entire AVF2119g study (Table 1), so the analysis of the minimum SLD% would lead to a slight reduction in specificity for this study. In study AVF2107g, the PFS and modeling analyses performed similarly, with the minimum SLD% performing the worst. Across the other 4 studies, the PFS comparison is either clearly the best (AVF2192g and SO14999) or among the best methods (BR21 and H0648g). Further, in these 4 studies, the minimum SLD% is better than the model-based analyses, although the difference is not great in SO14999.

## Discussion

Other tests for treatment effect evaluated, but not presented here, were the following:

- A Wald test was conducted of equality between the growth parameters in Model [1] with shrinkage parameters allowed to differ between treatment arms.

- Shrinkage and growth term estimates were obtained for each patient in each simulated Phase II trial via a simple nonlinear regression model (16–18) of SLD% on  $\exp(-St) + \exp(Gt) - 1$ . The growth terms were compared between treatment groups with a Wilcoxon rank-sum test.

These 2 tests tended to perform worse than the 2 modeling approaches presented here.

Other versions of Model [1] have been proposed. For example, ref. 12, the growth term can appear linearly as  $G_{ij} t_{ijk}$  instead of  $\exp(G_{ij} t_{ijk}) - 1$ . Because the linear term is the first-order Taylor series approximation to the exponential term, these 2 models would be expected to perform similarly for data, such as here, where the tumor assessments stop at patient progression.

It is surprising that the modeling approach is as good as a PFS analysis in study AVF2107g but has power only equal to or slightly greater than a Type I error rate in study AVF2192g. The analysis of SLD would tend to improve with more tumor assessments per patient, but AVF2192g was not an outlier in this regard. Specifically, the mean number of tumor assessments in the simulated trials from AVF2192g was 4.2, between the mean values of 3.0 assessments for BR21 and 5.3 assessments for AVF2107g. From Fig. 1, it does appear that the agreement between actual and predicted SLD is the worst in study AVF2192g. However, variability in tumor assessment would also be expected to affect the quality of PFS. In the end, no reason for the poor performance of the modeling approach in AVF2192g was found.

One possible criticism of the assessment here is that the overall PFS analysis was positive for 5 of the 6 studies evaluated, leading to a bias against the SLD analysis. However, these studies were not selected among many studies because of the strength of their PFS results, but because of the ready availability of the per-patient tumor assessments. In spite of the positive PFS results, there was nothing limiting the SLD analyses to be even better.

The most common reasons for patient exclusion from the analyses here (see Table 1) are being randomized but not treated, having nonmeasurable disease at baseline, and having no PFS value and no postbaseline tumor assessment data. These are patients who would reasonably be excluded from a Phase II study analysis. Another reason for patient exclusion, accounting for approximately 40% of excluded patients, is having a PFS value but no postbaseline tumor assessment data. It is difficult to glean the reasons for these outcomes from the databases, but frequently a new lesion was noted, progressive disease recorded, and target lesion measurements left missing. This highlights the importance, when SLD is a key study endpoint, of obtaining complete lesion assessments through progression. Regardless, a repeat of the PFS versus minimum SLD% analysis as in Fig. 2 with the inclusion of these 40% of excluded patients led to little difference in the results.

Although modeling of tumor burden appears not to be as useful as a PFS analysis for the decision to start a Phase

III trial, modeling has been used for other purposes. It would be interesting to know how a corresponding PFS analysis for these uses might fare in comparison. For example, the model-predicted SLD% at 8 weeks was used along with other patient characteristics to predict overall survival (OS) for second-line non-small cell lung carcinoma (NSCLC) patients (12). However, PFS could also be used in a model to predict OS, and an approach with an independent postprogression survival term added to PFS (30) might be competitive. Modeling was used to evaluate dose dependence of the growth parameter in a study where dose reductions were applied in some patients (15). PFS could also have been used to evaluate dose dependence by conducting a proportional hazards regression of PFS with recent dose as a time-dependent covariate.

The estimates of the individual patient growth terms were shown to be related to overall survival in renal cell carcinoma (16) and in breast cancer (18). For the 6 studies evaluated here, PFS and its censoring indicator were jointly stronger predictors of overall survival than were the shrinkage and growth parameter estimates from the simple nonlinear regression of SLD% on  $\exp(-St) + \exp(Gt) - 1$  (details not presented).

A potential advantage of an SLD analysis is that it can be conducted at an early assessment, leading to a faster decision whether to start Phase III. However, enrollment into Phase II trials takes time, and by the time the early assessment is available in the last patient enrolled, PFS will have been determined for many of the patients enrolled early, largely eliminating this potential advantage (19).

One reason why PFS appears to be the better endpoint than SLD in the evaluation of a Phase II trial may be that a patient can progress because of growth of nontarget lesions or the appearance of new lesions (31). These latter 2 reasons would not be reflected in changes in the SLD. Thus, PFS makes greater use of the information captured in the serial tumor assessments than does the SLD. There has been a recent suggestion (32) for a "longitudinal rank-based randomized phase II design, ranking a patient's risk of death, differentially weighting (disease progressions) by type and time of (progressive disease), and percentage change in tumor burden." The assessment of such an approach awaits further details on its implementation.

In conclusion, for the decision to start a Phase III trial based on the results of a randomized Phase II trial of an oncology drug, PFS appears from assessments performed to date to be the better endpoint than does SLD, whether analyzed through simple endpoints such as the minimum SLD% or through the modeling of its time course. It would be useful for these endpoints and analysis approaches to be assessed in further completed trials to achieve a more definitive overall conclusion or else to define those situations where an SLD analysis might be preferred.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.



## Acknowledgments

The author thanks the National Cancer Institute of Canada for permission to use the BR21 data.

Received June 28, 2012; revised October 17, 2012; accepted October 31, 2012; published OnlineFirst November 21, 2012.

## References

- Brown SR, Gregory WM, Twelves CJ, Buysse M, Collinson F, Parmar M, et al. Designing phase II trials in cancer: a systematic review and guidance. *Br J Cancer* 2011;105:194–199.
- Cannistra SA. Phase II trials in journal of clinical oncology. *J Clin Oncol* 2009;27:3073–3076.
- Ratain MJ, Sargent DJ. Optimising the design of phase II oncology trials: the importance of randomization. *Eur J Cancer* 2009;45:275–280.
- Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;23:7199–7206.
- Tang H, Foster NR, Grothey A, Ansell SM, Goldberg RM, Sargent DJ. Comparison of error rates in single-arm versus randomized phase II cancer clinical trials. *J Clin Oncol* 2010;28:1936–41.
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- Rubinstein L, Crowley J, Ivy P, LeBlanc M, Sargent D. Randomized phase II designs. *Clin Cancer Res* 2009;15:1883–90.
- Sharma MR, Stadler WM, Ratain MJ. Randomized phase II trials: a long-term investment with promising returns. *J Natl Cancer Inst* 2011;103:1093–1100.
- Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change of tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *J Natl Cancer Inst* 2007;99:1455–61.
- Ratain MJ, Eisen T, Stadler WM, Flaherty KT, Kaye SB, Rosner GL, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2006;24:2505–12.
- Buysse M, Quinaux E, Hendlitz A, Goulinopoulos V, Tournigand C, Mick R. Progression-free survival ratio as end point for phase II trials in advanced solid tumors. *J Clin Oncol* 2011;29:e451–e452.
- Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, et al. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin Pharmacol Ther* 2009;86:167–74.
- Claret L, Girard P, Hoff PM, Van Cutsem E, Zuideveld KP, Jorga K, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol* 2009;27:4103–8.
- Bruno R, Claret L. On the use of change in tumor size to predict survival in clinical oncology studies: towards a new paradigm to design and evaluate phase II studies. *Clin Pharmacol Ther* 2009;86:136–8.
- Stein A, Wang W, Carter A, Chiparus O, Hollaender N, Motzer R, et al. Dynamic tumor modelling of the RECORD-1 phase II trial of everolimus quantifies relationship between dose and tumor growth in metastatic renal cell carcinoma. *Eur Urol Suppl* 2011;10:232.
- Stein WD, Yang J, Bates SE, Fojo T. Bevacizumab reduces the growth rate constants of renal carcinomas: A novel algorithm suggests early discontinuation of bevacizumab resulted in a lack of survival advantage. *The Oncologist* 2008;13:1055–62.
- Stein WD, Huang H, Menefee M, Edgerly M, Kotz H, Dwyer A, et al. Other paradigms: growth rate constants and tumor burden determined using computed tomography data correlate strongly with the overall survival of patients with renal cell carcinoma. *Cancer J* 2009;15:441–7.
- Fojo AT, Stein WD, Wilkerson J, Bates SE. Kinetic analysis of breast tumor decay and growth following ixabepilone plus capecitabine (IXA + CAP) versus capecitabine alone (CAP) to discern whether the superiority of the combination is a result of slower growth, enhanced tumor cell kill, or both. *J Clin Onc* 28:15s, 2010 (suppl; abstr 1096).
- Fridlyand J, Kaiser LD, Fyfe G. Analysis of tumor burden versus progression-free survival for phase II decision making. *Contemp Clin Trials* 2011;32:446–52.
- Sharma MR, Karrison TG, Jin Y, Bies RR, Maitland ML, Stadler WM, et al. Resampling phase III data to assess phase II trial designs and endpoints. *Clin Cancer Res* 2012;18:2309–15.
- Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 2004;350:2335–42.
- Miller KD, Chap LI, Holmes FA, Cobleigh MA, Marcom PK, Fehrenbacher L, et al. Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *J Clin Oncol* 2005;23:792–9.
- Kabbinavar FF, Schulz J, McCleod M, Patel T, Hamm JT, Hecht JR, et al. Addition of bevacizumab to bolus fluorouracil and leucovorin in first-line metastatic colorectal cancer: Results of a randomized phase II trial. *J Clin Oncol* 2005;23:3697–705.
- O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub J-P, Cervantes G, et al. Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: Phase III trial results. *J Clin Oncol* 2002;20:2812–23.
- Shepherd FA, Pereira JR, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005;353:123–32.
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–92.
- Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207–14.
- Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Comput Stat Data Anal* 2005;49:1020–38.
- Monolix: A software for the analysis of nonlinear mixed effects models. [cited 2012 Jun 19]. Available from <http://www.lixoft.com/wp-content/resources/docs/UsersGuide.pdf>.
- Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 2009;101:1642–9.
- Rubinstein LV, Dancy JE, Korn EL, Smith MA, Wright JJ. Early average change in tumor size in a phase 2 trial: efficient endpoint or false promise? *J Natl Cancer Inst* 2007;99:1422–3.
- Mietlowski WL, Bao W, Wood PA, Williams DE, El-Hashimy M, Sarr C, et al. Clinical importance of including new and nontarget lesion assessment of disease progression (PD) to predict overall survival (OS): implications for randomized phase II study design. *J Clin Oncol* 30:15s, 2012 (suppl; abstr 2543).