

Enhanced Detection of Treatment Effects on Metastatic Colorectal Cancer with Volumetric CT Measurements for Tumor Burden Growth Rate Evaluation



Michael L. Maitland^{1,2}, Julia Wilkerson³, Sanja Karovic¹, Binsheng Zhao⁴, Jessica Flynn⁵, Mengxi Zhou³, Patrick Hilden⁵, Firas S. Ahmed⁴, Laurent Dercle⁴, Chaya S. Moskowitz⁵, Ying Tang⁶, Dana E. Connors⁷, Stacey J. Adam⁷, Gary Kelloff⁷, Mithat Gonen⁵, Tito Fojo³, Lawrence H. Schwartz⁴, and Geoffrey R. Oxnard⁸

ABSTRACT

Purpose: Mathematical models combined with new imaging technologies could improve clinical oncology studies. To improve detection of therapeutic effect in patients with cancer, we assessed volumetric measurement of target lesions to estimate the rates of exponential tumor growth and regression as treatment is administered.

Experimental Design: Two completed phase III trials were studied (988 patients) of aflibercept or panitumumab added to standard chemotherapy for advanced colorectal cancer. Retrospectively, radiologists performed semiautomated measurements of all metastatic lesions on CT images. Using exponential growth modeling, tumor regression (d) and growth (g) rates were estimated for each patient's unidimensional and volumetric measurements.

Results: Exponential growth modeling of volumetric measurements detected different empiric mechanisms of effect for each

drug: panitumumab marginally augmented the decay rate [tumor half-life; d [IQR]: 36.5 days (56.3, 29.0)] of chemotherapy [d : 44.5 days (67.2, 32.1), two-sided Wilcoxon $P = 0.016$], whereas aflibercept more significantly slowed the growth rate [doubling time; $g = 300.8$ days (154.0, 572.3)] compared with chemotherapy alone [$g = 155.9$ days (82.2, 347.0), $P \leq 0.0001$]. An association of g with overall survival (OS) was observed. Simulating clinical trials using volumetric or unidimensional tumor measurements, fewer patients were required to detect a treatment effect using a volumetric measurement-based strategy (32–60 patients) than for unidimensional measurement-based strategies (124–184 patients).

Conclusions: Combined tumor volume measurement and estimation of tumor regression and growth rate has potential to enhance assessment of treatment effects in clinical studies of colorectal cancer that would not be achieved with conventional, RECIST-based unidimensional measurements.

Introduction

For nearly 20 years, Response Evaluation Criteria in Solid Tumors (RECIST) has been the international consensus, structured, standardized method by which to evaluate new cancer therapeutics (1) in clinical studies. RECIST was developed and has been demonstrated to ensure robust, reproducible detection of treatment effects in patients with solid tumors across numerous clinical investigation sites with standard assessment methodology and technological resources, while also allowing comparisons with historical trials assessed using similar

methods (2). Despite the identification of numerous potential shortcomings of RECIST, (3–6), no superior alternative method for clinical trial analysis has been established (7).

CT scans are the most widely distributed and commonly used technology for multicenter clinical trials that employ RECIST. Various advances in CT image acquisition and analysis have also provided the potential to improve evaluation of diverse treatments for cancer. For example, with computer-aided segmentation algorithms, a radiologist can quickly define the boundaries of individual target lesions and efficiently obtain relatively precise volume measurements (8, 9). Such software for volumetric measurement has been developed and tested for about a decade. The methodology is scalable but, as a single technological advance to provide better outcomes correlation, tumor burden by volumetric measurement at the typical fixed time-points in clinical trials has offered modest advantage to assess treatment effects over the unidimensional measurements collected for RECIST-driven metrics such as progression-free survival (10–13).

Computational modeling of changes in serially measured tumor burden is well established as a method to assess treatment effects in animal models (14). Various computational methods to compare changes in tumor burden over time among treatment arms in human clinical trials have been proposed as an alternative to RECIST-based clinical trial metrics (15–18). Initial proposals for phase II trial designs suggested the change in tumor burden from baseline to the first assessment might be superior to RECIST-based objective response rates or progression-free survival (PFS). However, thoughtful critiques and sampling exercises suggested the simple conversion from the categorical and time-to-event assessments of RECIST to comparison of changes in the sum of the longest dimensions of all target lesions

¹Inova Schar Cancer Institute, Fairfax, Virginia. ²University of Virginia Cancer Center and Department of Medicine, Charlottesville, Virginia. ³Columbia University Herbert Irving Comprehensive Cancer Center, New York, New York. ⁴Department of Radiology, Columbia University Vagelos College of Physicians and Surgeons/New York Presbyterian Hospital, New York, New York. ⁵Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, New York. ⁶CCS Associates, McLean, Virginia. ⁷Foundation for the National Institutes of Health Biomarkers Consortium, North Bethesda, Maryland. ⁸Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Michael L. Maitland, University of Virginia Cancer Center, 8081 Innovation Park Drive, Third Floor, Fairfax, VA 22031. Phone: 571-472-1070; Fax: 571-472-1075; E-mail: michael.maitland@inova.org

Clin Cancer Res 2020;26:6464–74

doi: 10.1158/1078-0432.CCR-20-1493

©2020 American Association for Cancer Research.

Translational Relevance

Response Evaluation Criteria in Solid Tumors (RECIST) is often criticized, and new metrics based on methods of modeling the longitudinal growth of solid tumors have been suggested but have not been demonstrated superior to conventional RECIST metrics. Similarly, digital imaging techniques such as volumetric measurement of target lesions have been suggested to improve sensitivity and specificity for changes in tumor burden due to treatment, but the techniques have not been widely adopted for this purpose. In this retrospective analysis of completed colorectal cancer clinical trials, the largest to date of re-measurement of original CT images, we found potentially important advantages of combining a model-based metric with volumetric assessment of tumor burden. Our findings suggest that comparison of treatments based on calculation of the tumor growth rate, g , using volume measures of tumor directly on CT images achieves greater statistical power than using conventional unidimensional measurements.

alone would not offer substantial improvements (19–22) in statistical power or time-to-study-completion. In one of the largest examinations to date, Mandrekar and colleagues (23) examined the relationships among numerous RECIST-derived and continuous tumor burden metrics among more than 8,000 patients with either breast, colorectal, or lung cancer and overall survival (OS) captured in 13 clinical trials. They found no significant superiority of more complex methods over a simple trichotomous response metric at 24 weeks of treatment. Notably, the strongest relationship among imaged tumor burden metrics and OS was among patients with colorectal cancer. Without evidence of significantly improving performance of human clinical trials in solid tumors, volumetric measurement of tumor burden by CT imaging, and computational modeling of human tumor growth have gained interest but little traction in the clinical investigator community.

We hypothesized that when new computational modeling techniques of tumor growth inhibition are applied to better measurement of tumor burden on CT imaging, the combination of these two technologies would achieve the expected improvements in statistical power. As a combined strategy, tumor growth inhibition modeling and better methods of tumor burden assessment on CT images should be revisited. Global pharmaceutical companies have now effectively incorporated tumor growth inhibition models to inform oncology drug development decisions (24–26). Academic investigators have developed new methodologies to infer treatment effects from conventional imaging measurements (27–30). We have demonstrated that while RECIST is effective at reducing the “noisy” elements of conventionally measuring tumor burden and collecting data on clinical trials, the unfiltered measurement variance reduces the effectiveness of computational models (31). More recently, we demonstrated that when computed growth rate (g) for human prostate cancer is estimated from a biexponential model of serial quantitative measurement of serum prostate-specific antigen (PSA) there is a strong association between estimates of the rate of tumor growth, g , and OS (32). This relationship supports use of treatment-related changes in g as a clinical endpoint for early phase clinical trials to reduce the required sample size to detect life-prolonging effects of cancer treatment in human studies. But, to use g in this way for other solid tumors requires a robust, serially measurable quantitative measure of tumor burden.

Our research groups have joined the Foundation for NIH (FNIH) Biomarkers Consortium project “Advanced metrics and modeling with Volumetric CT for Precision Analysis of Clinical Trial Results (Vol-PACT)” (33). We hypothesized that for solid tumors without a robust serologic biomarker, computing metrics from measures derived directly from original study CT images could enhance detection of treatment effects. For this study we had access to CT images from completed phase III clinical trials that supported regulatory agency approval in colorectal cancer for the angiogenesis inhibitor, aflibercept, and the EGFR inhibitor, panitumumab. Notably these agents were added to and compared with a backbone of chemotherapy (panitumumab added to fluorouracil, folinic acid, oxaliplatin in first-line therapy, aflibercept added to fluorouracil, folinic acid, irinotecan in second-line therapy) and had modest impact on OS. Here we demonstrate: (i) the rate of growth of colorectal cancer, g , derived from CT images correlates well with OS, (ii) the biexponential model implies different mechanisms of drug effect (the VEGF inhibitor delayed tumor growth whereas the EGFR inhibitor enhanced the cytotoxic effects of chemotherapy), and (iii) the estimated statistical power to detect treatment effects in a clinical trial of colorectal cancer is enhanced to a greater extent when based on volumetric assessment of tumor burden and changes in growth rate, on routinely collected CT images, than when based on the RECIST-derived measurement of single longest dimensions.

Materials and Methods

Clinical trials and patients

Clinical trial source data were obtained and original CT images were subjected to standardized quality control procedures and analyzed through the FNIH Vol-PACT project as described previously (33). To support development of new phase II clinical trial study metrics and candidate endpoints for solid tumors, each clinical trial dataset in the Vol-PACT project was randomly divided into discovery/development and validation sets. In this study we focused specifically on patients from the CRC trials- “PRIME” (34) and “VELOUR” (35).

PRIME (34) was a randomized study of the EGFR inhibitor mAb panitumumab added to standard-of-care fluorouracil, folinic acid, oxaliplatin (FOLFOX4) in the first-line treatment of metastatic colorectal cancer. The study randomly assigned treatment for 1,183 patients. The prespecified analysis stratified patients by tumor *KRAS* codon 12 status and was performed on 93% of patients (1096), with 656 of these patients having tumors that did not bear a codon 12 mutation and were considered “wild-type” (WT). For this WT stratum, the HR for progression-free survival (PFS) was 0.80; 95% confidence interval (CI), 0.66–0.97; $P = 0.02$, favoring the panitumumab arm [median PFS was 9.6 months (95% CI, 9.2–11.1 months) for panitumumab-FOLFOX4 and 8.0 months (95% CI, 7.5–9.3 months) for FOLFOX4]. Median OS in the panitumumab arm was 23.9 months (95% CI, 20.3–28.3) and 19.7 months (95% CI, 17.6–22.6) for FOLFOX4 alone (HR = 0.83; 95% CI, 0.67–1.02; $P = 0.07$). CT imaging was performed pretreatment and every 8 weeks until progression. Patients were followed every 12 weeks for survival. In the original study, patients in the *KRAS* WT stratum had median follow up 13.2 months (range, 0–25.2 months) in the panitumumab arm and 12.5 months (range, 0–24.7 months) in the FOLFOX4 arm. For the Vol-PACT project, 626 of the original 656 WT patients had complete combined image and CRF data. These subjects were randomly allocated 2:1 to development and validation sets. For this investigation, we evaluated the 418 WT patients in the development set.

VELOUR was the registrational clinical trial of the VEGF-binding recombinant protein aflibercept or placebo added to fluorouracil, folinic acid, and irinotecan (FOLFIRI) in the second-line treatment of CRC (35). The study randomized 1226 patients and revealed a median improvement in OS for aflibercept (13.5 months) vs. placebo (12.1 months); HR = 0.82; CI, 0.71 to 0.94; $P = 0.0032$. Aflibercept also increased the progression-free survival (6.9 vs. 4.7 months; HR = 0.76; CI, 0.66 to 0.87; $P < 0.0001$) and the response rate 20% (CI, 16% to 23%) over placebo 11% (CI, 8.5% to 13.8%); $P = 0.0001$. CT imaging was performed prior to initiation of treatment and approximately every 6 weeks until disease progression. Patients were followed every 8 weeks after progression for survival. In the original trial, the reported median follow-up time for survival was 22.3 months. Of the 1,226 patients initially randomized to a treatment arm, 126 did not have data to pass Vol-PACT quality control. The remaining 1,140 with available clinical treatment and imaging data were randomized 1:1 to development/discovery and validation sets. In this investigation, we evaluated the 570 patients assigned to the development set.

Images and analysis procedures

Both studies had protocol-specified centralized CT imaging collection. The electronic transfer of the clinical and imaging data, the quality control procedures, import of images into the standard software platform, and segmentation analyses of lesions were described previously (33). Briefly, industry sponsor research teams recoded individual subject data and corresponding CT image files with the same subject identifier before transferring data and image files to the consortium. CT images were stored in DICOM format and transferred to the Columbia University Computational Image Analysis Lab (CIAL).

Tumor burden for each individual patient was assessed by a team of radiologists who were blinded to the associated clinical data using a response assessment system built on open source software, the Weasis imaging platform (8). Up to 10 lesions > 1 cm in diameter at baseline and new lesions upon appearance were segmented and measured at each scan time point. Measuring up to 10 lesions provided a better estimation of overall tumor burden than using the RECIST 1.1 rule of five maximum, and two per organ maximum. Semiautomated algorithms developed for lung lesions, liver lesions and lymph nodes with a contour modification tool were used. The contours were superimposed on the original images, reviewed by a radiologist, and corrected if deemed inaccurate by the reviewing radiologist. Once a lesion was segmented, its single longest dimension and volume could be calculated automatically by computer. The unidimension was the longest line length inside the segmented lesion (maximal diameter; in mm) calculated on the axial image (x - y plane) where the lesion has the largest area and volume was the total number of the voxels inside the segmented lesion multiplied by the image resolutions along x -, y -, and z -directions (voxel size; in mm^3).

Tumor growth modeling and notation

The same modeling method applied to analysis of sum of unidimensional measures of target lesions for renal cell carcinoma (36) and PSA for prostate cancer (32) was applied to the tumor burden assessment for patients with colorectal cancer. In addition to being a familiar method to clinicians, this method had the most readily available software package to apply to the dataset among published quantitative tumor growth inhibition models. The quantitative assessment for colorectal cancer was based on CT images for individual patients with colorectal cancer using either total unidimensional and volume measures. Conventionally, as per usual execution of RECIST-based trials, we have excluded from nonlinear model-based analysis, subjects with 2 or fewer CT imaging datasets, unless the difference in tumor measurements is clearly informative (between the single on-study and baseline image series is $\geq 20\%$, consistent with RECIST progressive disease criteria; ref. 37). Total cases excluded from subsequent modeling for this lack of informativeness is summarized in **Table 1** and specified by study arm in Supplementary Table S1. Individual cases are depicted in Supplementary Fig. S1. For all subjects with image collections meeting these criteria, the biexponential regression-growth model estimates the tumor growth rate based on the assumption that change in tumor quantity during therapy results from two independent component processes: an exponential decrease or regression, occurring at rate d , and a simultaneously occurring exponential growth or regrowth of the tumor, occurring at rate g . Consistent with prior evaluation and implementation of this methodology (36, 38–41) in thousands of patients, individual patient tumor burden trajectories by unidimensional and volumetric assessments of tumor burden conformed to four basic patterns related to the model:

- (i) When the biexponential relationship best fits the data (super majority of cases), the model is labeled gd , where $f(t)$ is the tumor quantity at time t in days, relative to initial tumor quantity, d is the rate of regression or decay, and g is the rate of growth:

$$f(t) = e^{-dt} + e^{(gt)}$$

- (ii) For patients in whom there is a continuous decrease in tumor burden from the start of treatment, the data are labeled dx , as the growth rate, g , is eliminated:

$$f(t) = e^{-dt}$$

- (iii) Similarly, d is eliminated when data show a continuous increase in tumor burden from the start of treatment, and labeled as gx :

$$f(t) = e^{gt}$$

- (iv) The fourth model, contains an additional parameter, ϕ , which represents the proportion of tumor cells that undergo cell death due to

Table 1. Total subjects with data available for analysis by study arm.

Study	Treatment arms	Measurement	Excluded <i>n</i> (%)	Analyzed fit <i>n</i> (%)	Analyzed not fit <i>n</i> (%)
VELOUR	FOLFIRI/placebo vs.	Unidimensional	94 (16.7)	445 (77)	31 (6.3)
	FOLFIRI/aflibercept	Volumetric	65 (11.4)	470 (81.7)	35 (6.9)
PRIME	FOLFOX vs.	Unidimensional	60 (16.3)	335 (77.3)	23 (6.4)
	FOLFOX/panitumumab	Volumetric	49 (11.7)	342 (80.4)	27 (7.9)

therapy. Sometimes, this more complex model better fits the data than gd and it is labeled as $gd\phi$:

$$f(t) = \phi e^{-dt} + (1 - \phi)e^{(gt)} - 1$$

Using all CT imaging measurements for unidimensional and volumetric assessments of tumor burden, the rates of tumor growth [g] and regression [d] were calculated to solve these four nonlinear least squares problems with the TUMGr package for R (32). The “initial tumor quantity” or “baseline tumor burden” was based on the investigator-determined CT imaging assessment consistent with the study protocol for each clinical trial. Also consistent with the protocols and conventional study execution, this session was almost always between days -28 to -1 of initial on-study treatment. Collected most closely of all measurements to study day 0, this measurement was assumed to be the effective measurement on the day of initial treatment, without adjustments for the model. For unidimensional measurements the units were in mm, and for volume measurements in mm^3 . For purposes of this modeling analysis of CT imaging-based lesions, we assumed the same measurement error for all diameters and volumetric assessments. Time was measured in days. Among models where all parameters are significant predictors ($P < 0.10$), the model which minimizes the Akaike Information Criterion is the selected model for a given patient. Typically, approximately 10% of patients who have sufficient imaging series and measurements have tumor burden data that do not fit any of the model structures well (Table 1, “Analyzed not fit”), that is, where no parameter predicts tumor burden with $P < 0.10$. In this study 6.3%–7.9% of patients per study arm had tumor burden data that did not fit the models. These subjects were excluded from subsequent survival and power analyses.

Survival analysis

To examine the association between g and OS, a landmark survival analysis was performed for each trial dataset and measurement type (unidimensional and volumetric) using a landmark time defined as the point where 75% of the measurement data had been collected for each trial, which were 10.1 and 5.9 months for PRIME and VELOUR trial sets, respectively. To prevent guarantee-time and immortal-time biases, only patients who lived to the landmark time point were included in the analysis. The log of the growth rate estimates obtained using measurement data prior to the landmark time was used as a single continuous predictor in the analyses. For visualization purposes, survival curves were depicted by tertile log g . The same approach was enlisted to evaluate the association between d and OS.

Power simulations

Power simulations for comparison of growth rates between experimental (aflibercept or panitumumab) and control (placebo) arms were performed for the volume and unidimensional data using two different methods: (i) a randomized study case example with prospective enrollment and assignment of patients to a treatment or control arm, and (ii) a single/historical comparator case example, where all patients would receive the investigational treatment and the difference in g would be compared with g for a similar population that has received the comparator treatment. In the randomized study case example for the given N , 1,000 random samples with replacement of size N were generated from the growth rates of the experimental and the control arm. For each of the 1,000 samples, a two-sided Wilcoxon test (null hypothesis: true location shift is equal to 0, and alternative

hypothesis: true location shift is not equal to 0) was performed against the control arm growth rates, and the test statistic P values were recorded. We chose Wilcoxon test for this comparison because of its well-known robustness against violations of assumptions involved with parametric methods like the t test on the original or log-transformed data. While there is a loss of statistical power associated with this choice, we found it important that our ability to control type I Error was not strongly dependent on the distribution of the estimated growth parameters. Power was then computed as the proportion of the 1,000 test statistics that were significant ($P \leq 0.05$). These steps were repeated for various values of N and the results were plotted (power $\sim N$), noting the N value at which a value of 0.80 for power was reached. For the single-arm/historical comparator simulations, for the given N , 1,000 random samples with replacement were generated from the experimental arm of size N . The analyses plotted in Fig. 4 assumed a $\log_{10} g$ value of -4 for all patients in whom the data best fit the dx model (where g cannot be estimated, because OS between the best g responses and the dx cases is the same and this replacement value preempts removal of these patients from the study (as would be desired in a prospective phase II investigation)). To determine the associated false-positive detection rate, we performed similar analyses but now resampled cohorts 20–100 patients each from the control arms of the PRIME and VELOUR studies and compared to the full control arm sample for 1,000 tests each and determined the frequency at which the study arm would be declared to improve g with $P < 0.05$.

Results

For these analyses, data were available from 988 patients with a diagnosis of colorectal cancer, 570 from VELOUR (phase III trial in second line treatment with FOLFIRI and placebo or FOLFIRI and aflibercept to patients whose disease had progressed on an oxaliplatin-based regimen), and 418 patients in the no-codon-12 *KRAS* mutation (WT) stratum of the PRIME trial (randomized phase III trial of first-line standard-of-care FOLFOX vs. FOLFOX and panitumumab).

All subject data were received from the set defined for the original study intention-to-treat analysis. The nonlinear analytical models best fit individual patient datasets with an available baseline scan and three or more tumor assessment time-points (Fig. 1). Models cannot be applied well to sets with fewer than three assessments. This nonvaluable fraction constituted 16%–17% of each of the trial datasets (Table 1). Use of volumetric measurement improved the evaluable fraction in both of the trials. Detailed evaluation of these exclusions of data from subsequent analysis (Supplementary Table S1) revealed no systematic differences among studies or treatment arms. The largest difference was between the placebo and aflibercept arms of VELOUR where the rate at which only two CT image series were available with $< 20\%$ change in measured tumor burden among subjects in the placebo arm (12%) was twice the rate of subjects in the aflibercept arm (6%). When volumetric measures rather than unidimensional measures were employed, the rates of exclusion decreased by roughly half (5% for the placebo arm and 3% for aflibercept).

Among the included cases, the individual patient data fits of the log ratio of tumor burden at each on-treatment time-point to baseline tumor burden over the course of study observations were directly inspected (Fig. 1) and fit one of the four model types: gd , dx , gx , and $gd\phi$. A consistent, but small, fraction of cases (5.9%–8.6%) did not meet minimum criteria for model fits and varied with study arm and whether unidimensional or volumetric assessments of tumor burden were employed (Supplementary Table S1).

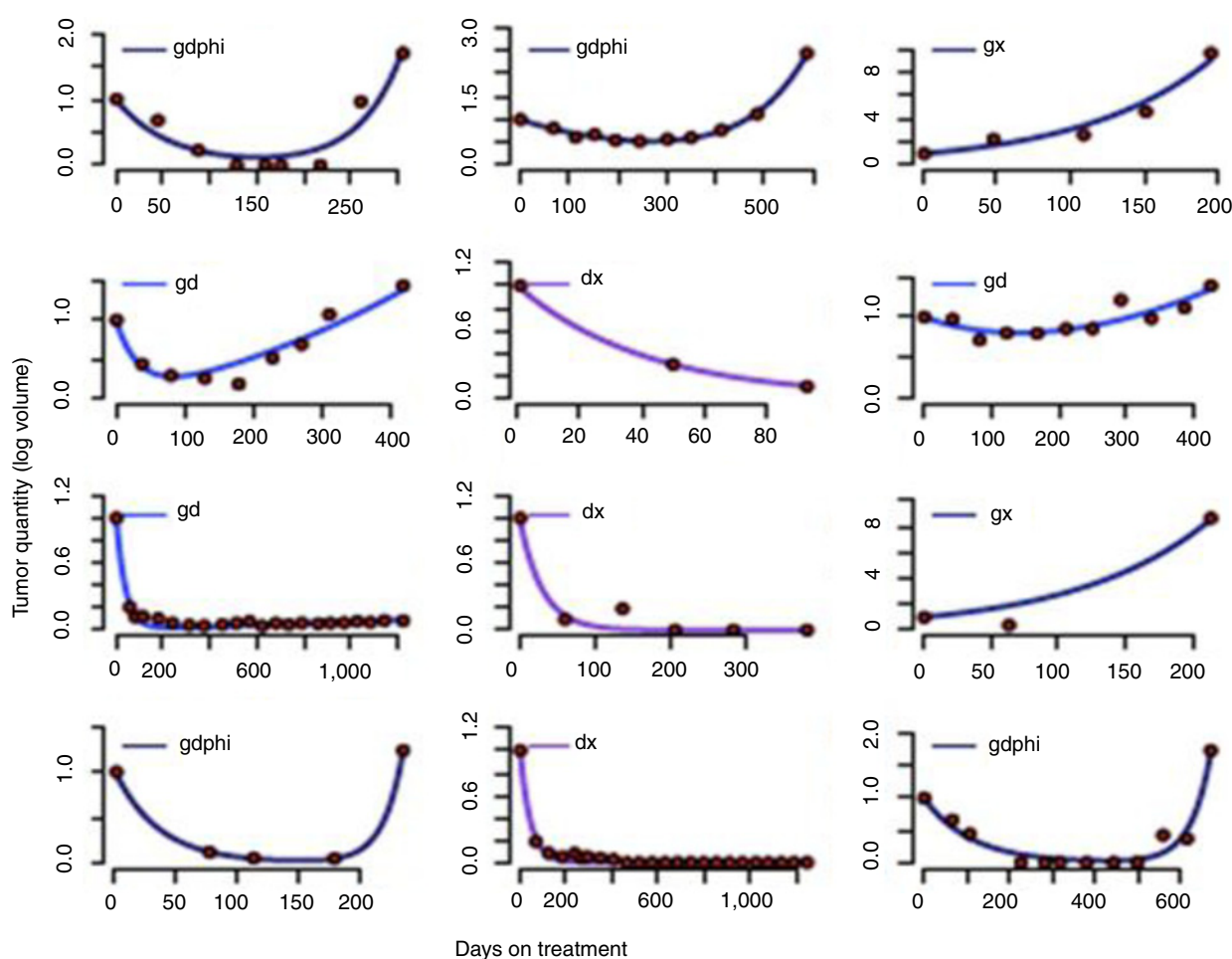


Figure 1.

Examples of curve fits. Shown are the observed radiographic values as red circles, and the best curve fit as a blue line. Note that the line is not drawn through the points. The shape of the line is described by the value of g and of d that is estimated from an analysis of the data. Each individual example is annotated with the equation that was used to fit the data. Curve fits for all patients analyzed can be found in Supplementary Fig. S1.

The estimates of g were lower in patients enrolled in the experimental arm of each trial, with greater differences in the VELOUR trial. As summarized in **Table 2** and shown graphically in **Fig. 2B** and **D**, the magnitude of the statistical difference was greater when comparing the volumetric data to the unidimensional data. This observation underscored the added value of volumetric measurements over unidimensional measures of tumor burden. This factor has important implications for design of future model-based early-phase clinical trials for colorectal cancer.

For the decay/regression constant, d , the analysis showed aflibercept to have no effect, but in PRIME a statistically significant difference was discernible with d higher with the addition of panitumumab, as shown in **Table 2** and Supplementary Fig. S2. This qualitatively different effect of aflibercept (primarily on g) and panitumumab (relatively more on d) implies that the different mechanisms of the drugs have different bases for augmenting chemotherapy effects on colorectal cancer. Panitumumab increases the initial reduction in tumor burden with marginal impact on the intrinsic growth rate of the tumor. In contrast, aflibercept has minimal impact on the chemotherapy-induced reduction in tumor burden and primarily slows the intrinsic growth rate of the tumor.

We have previously shown that g predicts OS in renal cell carcinoma (36) and prostate cancer (32) and in this analysis we have confirmed this relationship for colorectal cancer. **Figure 3** shows the data from both VELOUR and PRIME trials landmarked at the point when 75% of the data had been captured. Three curves depict tertiles of the values for g with the fourth curve representing the OS probabilities of the patients whose data was best fit by the dx model. These patients had no estimable g value, as this was either too small to estimate or nonexistent. The curve for these patients tracks closely with the curve for the best tertile of g values, consistent with a good OS probability for both groups. Notably, we confirmed for colorectal cancer in VELOUR and PRIME (Supplementary Fig. S3) our prior observations in renal cell carcinoma (36) and prostate cancer (41), that d does not correlate with OS (Supplementary Fig. S4).

Our estimates of changes in intrinsic growth, g , derived from volumetric measures of CT-imaged lesions in the VELOUR trial suggested potentially important increases in statistical power for testing new treatments in colorectal cancer. The rationale is as follows: (i) g has strong correlations with OS, (ii) aflibercept had measurable (although marginal) effects on OS, (iii) the difference in g between the placebo and aflibercept arms was more evident based on volumetric

Table 2. Comparison of parameter estimates by study arm with unidimensional versus volumetric measurement of target lesions on CT image series.

		<i>d</i> /half-life (days)							
VELOUR									
Analysis	Treatment	N	Mean	SD	Min	Q1	Median	Q3	Max
Unidimensional	FOLFIRI + placebo	144	247.7	390.6	27.2	99.0	145.0	222.7	3,758.3
	FOLFIRI + aflibercept	197	260.7	573.7	21.6	84.3	139.6	229.3	6,175.4
Volumetric	FOLFIRI + placebo	141	101.6	113.4	12.1	44.4	66.2	120.7	862.9
	FOLFIRI + aflibercept	189	94.9	102.3	13.2	43.3	71.2	110.1	1,095.8
		<i>g</i> /doubling time (days)							
VELOUR									
Analysis	Treatment	N	Mean	SD	Min	Q1	Median	Q3	Max
Unidimensional	FOLFOX	148	152.5	188.7	21.0	72.4	107.4	155.4	1,837.8
	FOLFOX + panitumumab	150	121.1	111.5	13.1	64.1	87.9	128.3	786.8
Volumetric	FOLFOX	140	66.7	106.2	12.6	32.1	44.5	67.2	1,073.0
	FOLFOX + panitumumab	151	63.7	160.7	11.4	29.0	36.5	56.3	1,948.0
		<i>g</i> /doubling time (days)							
VELOUR									
Analysis	Treatment	N	Mean	SD	Min	Q1	Median	Q3	Max
Unidimensional	FOLFIRI + placebo	198	743.0	4935.1	56.4	169.1	269.9	426.9	69,637.2
	FOLFIRI + aflibercept	190	413.8	347.5	59.8	219.4	311.8	505.1	3,380.1
Volumetric	FOLFIRI + placebo	204	589.5	3503.0	27.7	82.2	155.9	347.0	48,770.5
	FOLFIRI + aflibercept	195	527.4	853.9	20.9	154.0	300.8	572.3	9,505.7
		<i>g</i> /doubling time (days)							
PRIME									
Analysis	Treatment	N	Mean	SD	Min	Q1	Median	Q3	Max
Unidimensional	FOLFOX	140	904.0	1,552.9	28.0	250.1	434.1	1,011.3	11,729.0
	FOLFOX + panitumumab	136	1,173.8	2,250.8	60.0	336.9	591.0	1,025.6	16,914.9
Volumetric	FOLFOX	130	3,898.5	17,500.4	15.4	177.4	606.6	1,340.5	177,734.3
	FOLFOX + panitumumab	123	3,571.4	8,294.6	23.2	305.8	871.0	2,207.8	46,081.3

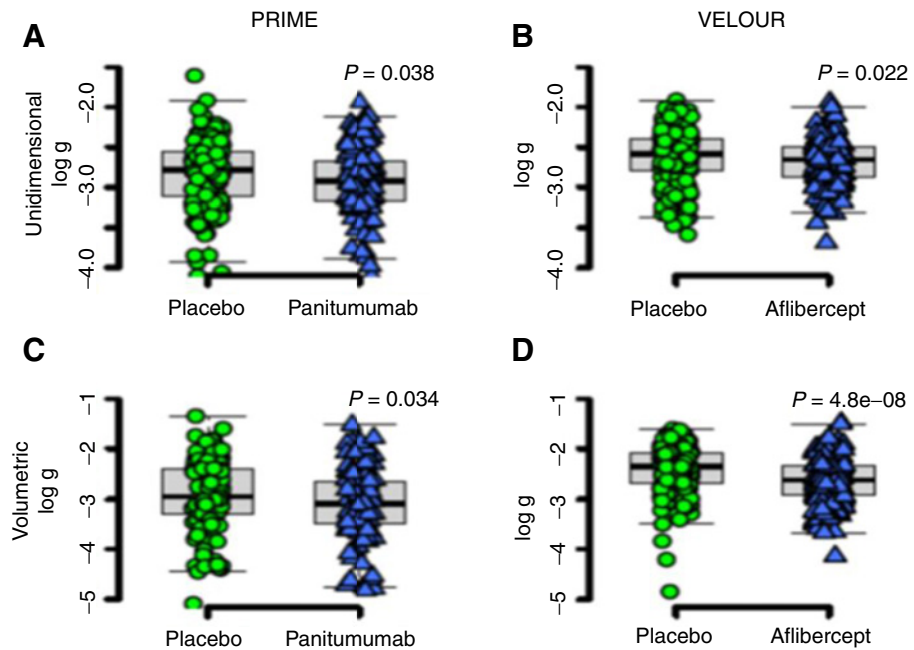
measures of tumor burden on CT than with unidimensional measures. We therefore explored the potential statistical power for an investigation based on detection of a magnitude of change in *g* that would clearly be suggestive of an improvement in OS.

We performed resampling analyses to determine the sample size required (Fig. 4) to detect the effects of the modeled treatment in

hypothetical phase II clinical trials based on growth modeling of tumor burden for patients in both arms of the trials with unidimensional versus volumetric measurements of tumor burden on CT images. We conceived two divergent study designs. In Fig. 4A and B, we evaluated the number of patients required to detect the effect of aflibercept added to FOLFIRI (Fig. 4A) or panitumumab added to FOLFOX (Fig. 4B) on

Figure 2.

Dot-plots of data in Table 2 comparing the distribution of *g* values in the two clinical trials derived from unidimensional [PRIME (A) and VELOUR (B)] or volumetric measures [PRIME (C) and VELOUR (D)]. In the PRIME trial, the difference was small magnitude and of marginal significance. In VELOUR a statistically significant difference was observed and the significance of this difference was amplified when based on the volumetric data.



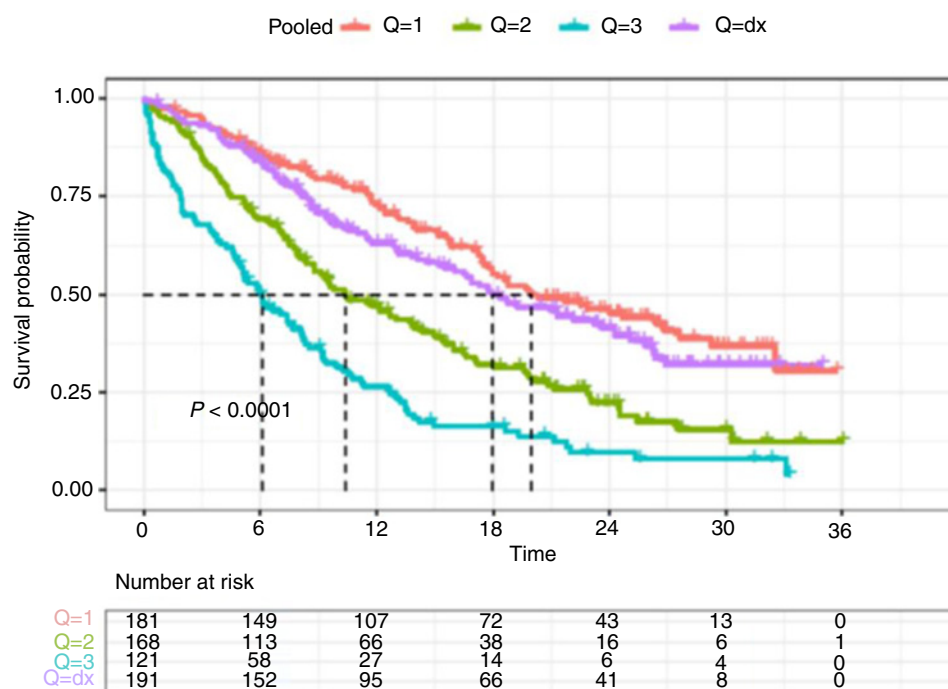


Figure 3.

Kaplan-Meier plots demonstrating the value of the growth rate constant, g , as a biomarker for overall survival. The predictive value was assessed by conducting a landmark analysis. Because data are from one first-line and one second-line trial the time at which to landmark the data was chosen as the time at which 75% of the data had been collected, rather than an arbitrary time in months. This was at 10.1 months for PRIME and 5.9 months for VELOUR. For this analysis, data from both VELOUR and PRIME were pooled together to demonstrate the value of g in this respect is indifferent to the trial from which it was gathered, and in this case to whether the trial was being conducted as first- or second-line therapy for metastatic colorectal cancer. The tertiles are of the patient data that had a calculable g within the landmark constraints and for which there was either a known OS value or evidence of when they were still alive allowing for censoring. The three tertiles have median OS values (95% CI) of 20 (17.7–27.6), 10.4 (8.8–13.9), and 6.1 (4.9–8.1) months, respectively, for those with the slowest g values, intermediate g values, and fastest g values from right to left, respectively. The fourth curve is that for patients whose data was best fit by the dx equation and in whom a g could not be calculated. Not surprisingly, these patients whose tumor either was not growing or growing very slowly have an OS (95% CI) 18 (15.5–24.3) months, comparable with those of the tertile with the lowest (slowest) g values.

g in a prospective, 1:1 randomized clinical trial. In a prospective study, we could not, *a priori*, identify patients whose patterns best fit the dx model, and so we substituted the same decrease in g , $\log_{10} -4$ as typically observed for the best tertile of patients, for the dx patients. To estimate the corresponding difference in g with 80% power in a one-sided test with α of 0.05 a conventional unidimensional-measurement-based study would require approximately 184 patients to detect the change in g . In contrast, the same study performed with volume measurements is predicted to require significantly fewer patients, approximately 60. The enhanced precision of tumor burden estimates over time could justify an alternative, single-arm trial-benchmarking approach where investigators could enlist a recent historical control arm to serve as comparator in a single-arm study of a new agent added to standard therapy (Fig. 4C and D). In this study design, (Fig. 4C) enlisting unidimensional measurement-based tumor burden in the growth model leads to a trial with 124 patients, but with volumetric measures, the same study could require as few as 32 patients to detect the addition of aflibercept as having a promising effect on colorectal cancer. Notably, in PRIME, panitumumab had minimal effects on g and so regardless of whether unidimensional or volumetric measurements are employed, the hypothetical phase II trials would require substantially more patients (Fig. 4B and D). In any of these cases, the spurious rate of detection of differences is small. We performed simulations with random sampling from each study control arm

compared with itself in cohorts of size 20–100 patients. For the FOLFOX arm of the PRIME trial, the rate of identifying improvement in g ($P < 0.05$) for the sampled cohort was 0.03, and for the FOLFIRI arm in the VELOUR trial the rate was 0.02. The rates were the same whether unidimensional or volumetric measurements were employed.

Discussion

To accelerate improvements in study of cancer therapies in humans, combinations of new technologies could be more effective than the individual processes. We evaluated the incorporation of direct measurements of tumor burden from CT images from nearly 1,000 patients enrolled in phase III clinical trials into a biexponential model of tumor growth inhibition. In this case, volumetric assessment improved statistical power to detect beneficial treatment effects over unidimensional measures. In addition, the direct measurement of tumor burden on CT images combined with biexponential modeling revealed differences in the effects of target-specific therapeutic proteins added to chemotherapy.

Increasingly, tumor growth inhibition modeling has become recognized as a powerful method to forecast outcomes, inform cancer drug development, and improve understanding of tumor dynamics in human studies (24, 26, 42, 43) However, modeling methods applied to

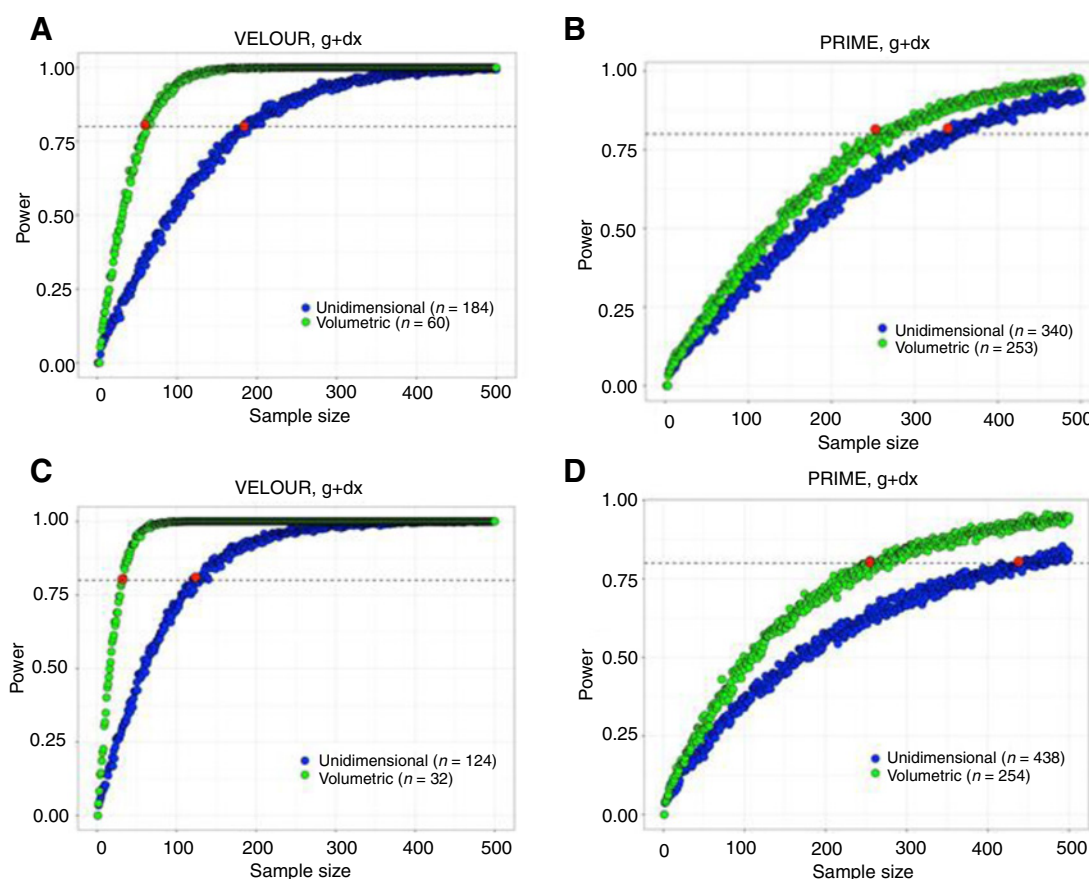


Figure 4.

Power simulation results for detection of drug effect on g with unidimensional versus volumetric methods of CT measurement by resampling data from (VELOUR trial with experimental aflibercept; **A** and **C**) and (PRIME trial with experimental panitumumab; **B** and **D**). Simulations entailed use of concurrent control arms of equivalent size (**A** and **B**) or consideration of larger, comparable historical control arms of larger size for benchmarking a single-arm study (**C** and **D**).

human subject CT imaging data derived from case report forms have not been widely accepted. In part, this is because metrics derived from imaging case report form–based methods of continuous measurement of tumor burden have not proved superior to RECIST as candidate endpoints for clinical trials. We have previously demonstrated that a strength of using categorical time-to-event metrics like progression-free survival based on RECIST measurements is that these assessments are robust to interobserver variability in measurement and other sources of “noise” in capturing tumor treatment effects on tumor burden. Simultaneously, we demonstrated that computational modeling of these unidimensional CT image measurement data can lead to error propagation and poor performance of a model in reflecting the effects of treatment (31). We also have demonstrated that with segmentation algorithm–based measurement of tumor volume directly on CT images, the precision of measurement relative to the changes in tumor size is superior to unidimensional measurements in colorectal cancer (9). Our resampling analysis of the VELOUR trial here suggests that the combination of improved precision of volume measurement and computational analysis of tumor growth and decay offers substantial potential to reduce sample sizes for phase II solid tumor clinical trials.

For volume measurement combined with biexponential modeling, our findings suggest basic mechanistic or biomarker effects could be

better detected in early-phase oncology clinical trials or small retrospective analyses. In this dataset we observed that the anti-EGFR mAb panitumumab, when added to first-line chemotherapy causes a small but measurable increase in the regression rate of colorectal cancer in patients without a *KRAS* codon 12/13 mutation, with a marginally detectable effect on the growth rate of the disease. Meanwhile, aflibercept had no detected effects on the regression rate but a larger effect on the intrinsic growth rate that was more readily detected with use of volume measures of tumor burden.

For this initial effort at modeling tumor burden measures directly from CT images, sponsors provided images and data from older clinical trials. Effects of the investigational agents were modest. Disappointingly, currently recognized valid prognostic and predictive covariates such as the side of origin of colorectal cancer were not collected in the original clinical trial databases. For PRIME and panitumumab knowledge of somatic mutation markers of EGFR inhibitor resistance has evolved since completion of the trial. For example, a follow-up analysis of available tissue from the trial revealed approximately 17% of patients in this “WT” cohort to have tumors bearing other-site *KRAS*, *BRAF*, or *NRAS* mutations which caused a similar degree of primary resistance to panitumumab as the codon 12/13 *KRAS* mutations (44). These data were generated after the completion of the original clinical trial and so were not available for incorporation into this analysis. Since then, additional somatic

mutations in other genes conferring resistance to EGFR inhibition have been described (45). The presence of primary right or left-side colon tumors was not determined in the original PRIME trial dataset, but an independent team conducted a retrospective chart analysis specific for the WT patient population (46). Meta-analyses of other trials (47) and including PRIME (48) consistently detect a predictive effect of tumor-sidedness on anti-EGFR therapy. The method of assessment of d and g as we've described would offer a means to detect objective differences in regression or growth rate by side of primary tumor and provide a mechanistic validation of the effect of EGFR antibodies on survival by analysis of a subset of CT images from a study of anti-EGFR therapy added to colorectal cancer. The larger VolPACT project has restricted 1/3 of the available data from PRIME (and 1/2 from VELOUR) for future validation studies, and after completion of specified pan-trial analyses in progress, we ultimately could pursue this analysis.

A shortcoming to this approach to detecting treatment effects is that not all enrolled patients will be evaluable. Rather than consider intention-to-treat, a study based on these methods would have to focus on patients who are informative- have measurable target lesions that can be identified on the baseline images and serially measured on at least two additional CT scans. The patients would need to have remained on assigned treatment over that interval and the measurement data would have to then fit the model employed here. Across VELOUR and PRIME by our analysis, approximately 15%–20% of patients were not informative and therefore excluded from the analysis. Therefore, our estimates of sample size for volume measurement-based trials described above for 32–60 patients with informative data would require actual initial accrual of 40–80 patients. This also presents challenges to using the combination of volume measurements and biexponential-modeling-based metrics as the basis for a new prospective clinical trial endpoint or for monitoring and making treatment-decisions on individual patients. However, to discern important differences in treatment effects among small groups of patients without objective responses, our study suggests there is potential for improved statistical power to explore for exposure/response and biomarker-based effects. New conditional survival model methods might also benefit from incorporation of volumetric assessments of tumor burden (30). Other published methods to model these data such as with the nonlinear mixed-effects approach could reduce some bias in estimates of “ d ” and “ g ” and possibly further enhance detection of treatment effects. Also, the use of newer machine-learning-based techniques is compelling. A future goal for VolPACT is to facilitate use of these alternative approaches.

To our knowledge, this is the largest study to date of tumor volume measurements directly from CT images in clinical trials. This multi-institutional effort reflects advances in collaborative operations among imagers, computational scientists, and clinicians that should facilitate more powerful investigations of new cancer treatments and biomarkers in subsets of patients through the combined efforts of investigators with complementary methodology and expertise. Our findings suggest that direct measurement of tumor burden from CT images, more so with volume than unidimensional measures leads to capacity to detect treatment effects that are associated with OS in dozens rather than hundreds of patients. This Foundation for the NIH Biomarkers Consortium VolPACT project has generated a repository of original clinical trial data and the CT image files to support intensive assessment of alternative strategies. The findings in this study suggest that the combination of direct measurement of lesions from images as a

reflection of tumor burden, combined with growth modeling could be an important advance over modeling of conventional RECIST-based unidimensional target lesions alone.

Disclosure of Potential Conflicts of Interest

M.L. Maitland reports a contract from Foundation for the NIH (/contract) and grants from NCI during the conduct of the study, and M.L. Maitland's spouse is a cardiologist/clinical epidemiologist who is routinely consulted by biotechnology and pharmaceutical companies on development of new treatments for pulmonary hypertension and right ventricular heart failure; during the past three years, the only sponsor with overlapping interests between her unrelated work and this manuscript is Merck, Sharp, and Dohme. L.H. Schwartz reports grants from Merck (member image endpoint committee), personal fees from Regeneron (data safety and endpoint committee member), and personal fees from Boehringer (data safety and endpoint committee member) outside the submitted work. G.R. Oxnard reports personal fees and other from Foundation Medicine (employment) outside the submitted work. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

M.L. Maitland: Conceptualization, resources, data curation, supervision, funding acquisition, validation, investigation, methodology, writing-original draft, project administration, writing-review and editing. **J. Wilkerson:** Conceptualization, formal analysis, investigation, methodology, writing-original draft. **S. Karovic:** Data curation, writing-original draft, writing-review and editing. **B. Zhao:** Conceptualization, resources, data curation, formal analysis, methodology, project administration, writing-review and editing. **J. Flynn:** Data curation, formal analysis, methodology, writing-original draft. **M. Zhou:** Data curation, formal analysis, validation, writing-review and editing. **P. Hilden:** Data curation, formal analysis. **F.S. Ahmed:** Methodology, writing-review and editing. **L. Derclé:** Methodology, writing-review and editing. **C.S. Moskowitz:** Conceptualization, data curation, supervision, validation, methodology, project administration, writing-review and editing. **Y. Tang:** Conceptualization, data curation, project administration, writing-review and editing. **D.E. Connors:** Conceptualization, resources, funding acquisition, project administration, writing-review and editing. **S.J. Adam:** Conceptualization, resources, data curation, supervision, funding acquisition, project administration, writing-review and editing. **G. Kelloff:** Conceptualization, resources, supervision, funding acquisition, writing-review and editing. **M. Gonen:** Conceptualization, resources, formal analysis, supervision, validation, methodology, project administration, writing-review and editing. **T. Fojo:** Conceptualization, resources, data curation, software, formal analysis, supervision, methodology, writing-original draft, project administration. **L.H. Schwartz:** Conceptualization, resources, data curation, formal analysis, supervision, methodology, project administration, writing-review and editing. **G.R. Oxnard:** Conceptualization, resources, supervision, funding acquisition, methodology, writing-original draft, project administration.

Acknowledgments

Scientific and financial support for the Foundation for the National Institutes of Health Biomarkers Consortium project Vol-PACT (Advanced Metrics and Modeling with Volumetric Computed Tomography for Precision Analysis of Clinical Trial Results) was provided by: Amgen; Boehringer Ingelheim; Merck KGaA, Darmstadt, Germany; Genentech; Merck & Co., Inc.; Regeneron Pharmaceuticals; and Takeda Pharmaceutical Company. In-kind donations of phase III trial data to support this specific study were provided to Foundation for the National Institutes of Health by Amgen and Sanofi. Additional support was provided by: NIH R01-CA194783 (M.L. Maitland, S. Karovic, B. Zhao, L.H. Schwartz), 1U01-CA225431 (to B. Zhao and L.H. Schwartz), P30 CA008748 (J. Flynn, C.S. Moskowitz, M. Gonen).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 30, 2020; revised August 2, 2020; accepted September 23, 2020; published first September 28, 2020.

References

- Litiere S, Isaac G, De Vries EGE, Bogaerts J, Chen A, Dancy J, 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* 2019;37:1102–10.
- Oxnard GR, Morris MJ, Hodi FS, Baker LH, Kris MG, Venook AP, et al. When progressive disease does not mean treatment failure: reconsidering the criteria for progression. *J Natl Cancer Inst* 2012;104:1534–41.
- Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. *J Clin Oncol* 2004;22:4442–5.
- Sharma MR, Maitland ML, Ratain MJ. RECIST: no longer the sharpest tool in the oncology clinical trials toolbox—point. *Cancer Res* 2012;72:5145–9.
- Benjamin RS, Choi H, Macapinlac HA, Burgess MA, Patel SR, Chen LL, et al. We should desist using RECIST, at least in GIST. *J Clin Oncol* 2007;25:1760–4.
- Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009;15:7412–20.
- Fojo AT, Noonan A. Why RECIST works and why it should stay—counterpoint. *Cancer Res* 2012;72:5151–7.
- Yang H, Schwartz LH, Zhao B. A response assessment platform for development and validation of imaging biomarkers in oncology. *Tomography* 2016;2:406–10.
- Zhao B, Lee SM, Lee HJ, Tan Y, Qi J, Persigehl T, et al. Variability in assessing treatment response: metastatic colorectal cancer as a paradigm. *Clin Cancer Res* 2014;20:3560–8.
- Dicken V, Bornemann L, Moltz JH, Peitgen HO, Zaim S, Scheuring U. Comparison of volumetric and linear serial CT assessments of lung metastases in renal cell carcinoma patients in a clinical phase IIB study. *Acad Radiol* 2015;22:619–25.
- Mozley PD, Bendtsen C, Zhao B, Schwartz LH, Thorn M, Rong Y, et al. Measurement of tumor volumes improves RECIST-based response assessments in advanced lung cancer. *Transl Oncol* 2012;5:19–25.
- Mozley PD, Schwartz LH, Bendtsen C, Zhao B, Petrick N, Buckler AJ. Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence. *Ann Oncol* 2010;21:1751–5.
- Wulff AM, Fabel M, Freitag-Wolf S, Tepper M, Knabe HM, Schafer JP, et al. Volumetric response classification in metastatic solid tumors on MSCT: initial results in a whole-body setting. *Eur J Radiol* 2013;82:e567–73.
- Looney WB, Trefil JS, Schaffner JG, Kovacs CJ, Hopkins HA. Solid tumor models for the assessment of different treatment modalities: systematics of response to radiotherapy and chemotherapy. *Proc Natl Acad Sci U S A* 1976;73:818–22.
- Lavin PT. An alternative model for the evaluation of antitumor activity. *Cancer Clin Trials* 1981;4:451–7.
- Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *J Natl Cancer Inst* 2007;99:1455–61.
- 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.
- 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.
- 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.
- 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.
- An MW, Dong X, Meyers J, Han Y, Grothey A, Bogaerts J, et al. Evaluating continuous tumor measurement-based metrics as phase II endpoints for predicting overall survival. *J Natl Cancer Inst* 2015;107:djv239.
- Kaiser LD. Tumor burden modeling versus progression-free survival for phase II decision making. *Clin Cancer Res* 2013;19:314–9.
- Mandrekar SJ, An MW, Meyers J, Grothey A, Bogaerts J, Sargent DJ. Evaluation of alternate categorical tumor metrics and cut points for response categorization using the RECIST 1.1 data warehouse. *J Clin Oncol* 2014;32:841–50.
- Chatterjee MS, Elassaiss-Schaap J, Lindauer A, Turner DC, Sostelly A, Freshwater T, et al. Population pharmacokinetic/pharmacodynamic modeling of tumor size dynamics in pembrolizumab-treated advanced melanoma. *CPT Pharmacometrics Syst Pharmacol* 2017;6:29–39.
- Chigutsa E, Long AJ, Wallin JE. Exposure-response analysis of necitumumab efficacy in squamous non-small cell lung cancer patients. *CPT Pharmacometrics Syst Pharmacol* 2017;6:560–8.
- Claret L, Jin JY, Ferte C, Winter H, Girish S, Strohm M, et al. A model of overall survival predicts treatment outcomes with atezolizumab versus chemotherapy in non-small cell lung cancer based on early tumor kinetics. *Clin Cancer Res* 2018;24:3292–8.
- Ferte C, Fernandez M, Hollebecque A, Koscielny S, Levy A, Massard C, et al. Tumor growth rate is an early indicator of antitumor drug activity in phase I clinical trials. *Clin Cancer Res* 2014;20:246–52.
- Champiat S, Dercle L, Ammari S, Massard C, Hollebecque A, Postel-Vinay S, et al. Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti-PD-1/PD-L1. *Clin Cancer Res* 2017;23:1920–8.
- Rayfield CA, Grady F, De Leon G, Rockne R, Carrasco E, Jackson P, et al. Distinct phenotypic clusters of glioblastoma growth and response kinetics predict survival. *JCO Clin Cancer Inform* 2018;2:1–14.
- Tardivon C, Desmées S, Keroui M, Bruno R, Wu B, Mentre F, 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* 2019;106:810–20.
- Li CH, Bies RR, Wang Y, Sharma MR, Karovic S, Werk L, et al. Comparative effects of CT imaging measurement on RECIST end points and tumor growth kinetics modeling. *Clin Transl Oncol* 2016;9:43–50.
- Wilkerson J, Abdallah K, Hugh-Jones C, Curt G, Rothenberg M, Simantov R, et al. Estimation of tumour regression and growth rates during treatment in patients with advanced prostate cancer: a retrospective analysis. *Lancet Oncol* 2017;18:143–54.
- Dercle L, Connors DE, Tang Y, Adam SJ, Gonen M, Hilden P, et al. Vol-PACT: a foundation for the NIH public-private partnership that supports sharing of clinical trial data for the development of improved imaging biomarkers in oncology. *JCO Clin Cancer Inform* 2018;2:1–12.
- Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol* 2010;28:4697–705.
- Van Cutsem E, Tabernero J, Lakomy R, Prenen H, Prausova J, Macarulla T, et al. Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J Clin Oncol* 2012;30:3499–506.
- Stein WD, Wilkerson J, Kim ST, Huang X, Motzer RJ, Fojo AT, et al. Analyzing the pivotal trial that compared sunitinib and IFN-alpha in renal cell carcinoma, using a method that assesses tumor regression and growth. *Clin Cancer Res* 2012;18:2374–81.
- 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.
- Stein WD, Figg WD, Dahut W, Stein AD, Hoshen MB, Price D, et al. Tumor growth rates derived from data for patients in a clinical trial correlate strongly with patient survival: a novel strategy for evaluation of clinical trial data. *Oncologist* 2008;13:1046–54.
- 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. *Oncologist* 2008;13:1055–62.
- Stein WD, Huang H, Menefee M, Ederly 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.
- Stein WD, Gully JL, Schlom J, Madan RA, Dahut W, Figg WD, et al. Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. *Clin Cancer Res* 2011;17:907–17.
- Khan KH, Cunningham D, Werner B, Vlachogiannis G, Spiteri I, Heide T, et al. Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial. *Cancer Discov* 2018;8:1270–85.

43. Bruno R, Bottino D, de Alwis DP, Fojo AT, Guedj J, Liu C, et al. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. *Clin Cancer Res* 2020;26:1787–95.
44. Douillard JY, Oliner KS, Siena S, Taberero J, Burkes R, Barugel M, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* 2013;369:1023–34.
45. Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* 2015;526:263–7.
46. Boeckx N, Koukakis R, Op de Beeck K, Rolfo C, Van Camp G, Siena S, et al. Primary tumor sidedness has an impact on prognosis and treatment outcome in metastatic colorectal cancer: results from two randomized first-line panitumumab studies. *Ann Oncol* 2017;28:1862–8.
47. Tejpar S, Stintzing S, Ciardiello F, Taberero J, Van Cutsem E, Beier F, et al. Prognostic and predictive relevance of primary tumor location in patients with RAS wild-type metastatic colorectal cancer: retrospective analyses of the CRYSTAL and FIRE-3 trials. *JAMA Oncol* 2017;3:194–201.
48. Arnold D, Lueza B, Douillard JY, Peeters M, Lenz HJ, Venook A, et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. *Ann Oncol* 2017;28:1713–29.