Analyzing the Pivotal Trial That Compared Sunitinib and IFN-α in Renal Cell Carcinoma, Using a Method That Assesses Tumor Regression and Growth

Wilfred D. Stein1,2, Julia Wilkerson1, Sindy T. Kim3, Xin Huang3, Robert J. Motzer4, Antonio Tito Fojo1, and Susan E. Bates1

Abstract

Purpose: We applied a method that analyzes tumor response, quantifying the rates of tumor growth (g) and regression (d), using tumor measurements obtained while patients receive therapy. We used data from the phase III trial comparing sunitinib and IFN-α in metastatic renal cell carcinoma (mRCC) patients.

Methods: The analysis used an equation that extracts d and g.

Results: For sunitinib, overall survival (OS) was strongly correlated with log g (Rsq = 0.44, P < 0.0001); much less with log d (Rsq = 0.04; P = 0.0002). The median g of tumors in these patients (0.00082 per days; log g = -3.09) was about half that (P < 0.001) of tumors in patients receiving IFN-α (0.0015 per day; log g = -2.81). With IFN-α, the OS/log g correlation (Rsq = 0.14) was weaker. Values of g from measurements obtained by study investigators or central review were highly correlated (Rsq = 0.80). No advantage resulted in including data from central review in regressions. Furthermore, g can be estimated accurately four months before treatment discontinuation. Extrapolating g in a model that incorporates survival generates the hypothesis that g increased after discontinuation of sunitinib but did not accelerate.

Conclusions: In patients with mRCC, sunitinib reduced tumor growth rate, g, more than did IFN-α. Correlating g with OS confirms earlier analyses suggesting g may be an important clinical trial endpoint, to be explored prospectively and in individual patients. Clin Cancer Res; 18(8); 2374–81. ©2012 AACR.

Introduction

In 2010, approximately 570,000 people died of cancer in the United States, most from chemotherapy-refractory solid tumors, including more than 13,000 from metastatic renal cell carcinoma (mRCC; ref. 1). With the advent of sorafenib, sunitinib, temsirolimus, bevacizumab with IFN, everolimus, pazopanib, and now axitinib therapy of mRCC has improved (2–7). However, these therapies are not curative, underscoring the need for alternative treatment strategies and novel decision paradigms (8).

We developed a method to analyze tumor responses to therapy, quantifying the rate of tumor regression (decay, d) and growth (g), using measurements obtained while patients receive therapy (9–12). The rate constants are derived using data collected in clinical trials. In responding tumors, regression dominates from start of therapy until nadir, growth dominating after nadir. If tumors do not respond, growth dominates throughout. Previous analyses of phase II studies in mRCC used computed tomography (CT) measurements of tumors (9, 11); in prostate cancer, serum prostate-specific antigen (10, 12) was used. For those, growth rate constants, g, correlated with overall survival (OS), while, surprisingly, regression rate constants, d, did not.

Here, we used randomized phase III trial data comparing sunitinib with IFN-α in untreated mRCC (3, 13). We illustrate the value of the growth rate constant as a clinical trial endpoint, reflecting the impact of sunitinib on tumor growth.

Materials and Methods

Clinical trial and study design

The study, an international, multicenter, randomized, phase III trial, compared sunitinib (SU11248, Pfizer), with IFN-α (3, 13). OS was calculated from randomization date until date of death. Tumor measurements from CT scans were recorded as the sum of longest diameters (LD) of...
Translational Relevance

Sunitinib was the first agent in metastatic renal cell cancer to show major clinical responses in a significant subset of patients; the randomized trial against IFN made it the first-line choice. However, survival analysis was complicated by the use of sunitinib and similar agents after disease progression. The growth rate constant, \( g \), is a clinical trial endpoint that could have aided development of sunitinib by documenting reduced tumor growth rate early in the clinical trial. Results of the analysis of \( g \) are highly suggestive that reduced \( g \) can be equated with clinical benefit. A hypothesis-generating model incorporating \( g \) also suggests that continuing therapy after conventional measures of progression in a patient with marked reduction in \( g \) could increase survival. These analyses suggest that calculation of \( g \) can aid in both drug development and individual patient management.

Mathematical, data, and statistical analyses

Mathematical analysis. Our regression-growth equation is based on the assumption that change in tumor quantity during therapy, indicated by change in the sum of LDs, results from 2 independent component processes (both following first order kinetics): an exponential decrease/regression, \( d \), and an exponential growth/growth of the tumor, \( g \). The equation is:

\[
f(t) = \exp(-d \cdot t) + \exp(g \cdot t) - 1 \quad (A)
\]

where \( f(t) \) is the tumor quantity (sum of LDs) at time \( t \) in days, normalized to the sum of LDs at \( t = 0 \), and \( d \) (decay, fraction per day) and \( g \) (growth, also per day) are the pertinent rate constants; \( \exp \) is the base of the natural logarithms.

Theoretical curves depicting the separate components of Eq. (A) and how these combine together to give the time dependence of \( f \), the tumor quantity, appear in Fig. 1. For data showing a continuous decrease from start of treatment, \( g \) is eliminated:

\[
f(t) = \exp(-d \cdot t) \quad (B)
\]

Similarly, when tumor measurements show a continuous increase, \( d \) is eliminated:

\[
f(t) = \exp(g \cdot t) \quad (C)
\]

Data analysis. We attempted to fit each data set for which more than one data point was available. Curves were fit and parameters estimated with the procedure NLIN, the non-linear regression model, in SAS\(^1\) (1) software. The routine extracts parameters \( g \) and \( d \) with associated Student \( t \) and \( P \) values.

Statistical analyses. Linear regressions used the polynomial linear routine of SigmaPlot 11.0 (Systat Software). Sample comparisons of data sets were on SigmaPlot 11.0 with Student \( t \) test for normally distributed data, or the Mann–Whitney rank-sum procedure for data that were not normally distributed. We report the appropriate \( P \) values for a 2-sided assessment.

\(^1\)The output for this paper was generated using Base SAS software\(^6\), Version [9.1.3] of the SAS System for Windows. Copyright © [2002—2003] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks of SAS Institute Inc., Cary, NC, USA.
Results
Analyzing the time course of tumor decay and growth
The data analyzed were obtained in the phase III registration trial of patients with mRCC randomized to either sunitinib or IFN-α (3, 13). There were 379 patients assigned to sunitinib and 377 to IFN-α. Tumor measurements were assessed as the sum of the LDs of target lesions. Figure 1 depicts 4 examples of on-study change in LD sum, depicted as solid circles. Solid lines are curve-fits of the equations describing the sum of concomitant regression (decay) and growth fractions of the tumor, represented as dotted and dashed lines, respectively [Eqs. (A–C)]. Figure 1A and B show a decrease in LD sum followed by an increase. Figure IC depicts a case in which no growth is seen during treatment, whereas Fig. 1D shows a case with no apparent tumor shrinkage. We extracted g and d parameters, excluding patients with no tumor assessment data, with only baseline CT scan data, or with only one follow-up assessment that differed by less than 1.2-fold, these last patients not meeting criteria for progressive disease-treatment discontinued for toxicity or withdrawal of consent (3). Exclusions left 331 patients in the sunitinib arm and 268 in the IFN-α arm. A consort diagram showing patient numbers is in Supplementary Fig. S1. Setting the significance for the derived parameters at P < 0.1, we extracted acceptable g and/or d parameters in 319 (84%) of the patients assigned to sunitinib, and 240 (64%) of those assigned to IFN-α (differences in proportions, P = 0.007, χ² test).

To test whether the difference in ability to extract a g value indicated different patient populations, we compared OS, initial tumor quantity Q₀, nadir fraction, and time to nadir for patients with acceptable g/d parameters (i.e., extracted at P < 0.1) with those for the whole population (Supplementary Table S1). For both IFN-α and sunitinib arms, for either deceased patients or those still living at study closure, neither OS, initial tumor quantity, nadir fraction, nor time to nadir were different (every P > 0.25, Mann–Whitney rank-sum) between those with acceptable g/d and the whole population. In addition, Kaplan–Meier survival plots were not different between those with acceptable g/d parameters and those without (log rank: medians for sunitinib 130.6 and 122.7 weeks, P = 0.79, for IFN-α, 137.7, 119.4, P = 0.69). Thus, the cases with acceptable g/d parameters were representative of the whole population.

To validate the goodness of fit, we tabulated the R² values for the data points and fitted line for 20 cases selected randomly among those that had an acceptable g parameter. The median R² was 0.933, (25%–75%: 0.891–0.958), indicating acceptable agreement between fitting equation and tumor quantity values over the data set. In addition, the median of the P values for all acceptable g/d values was P = 0.00003 (25%–75%: 0.000–0.00192) for the sunitinib arm, and P = 0.000540, (25%–75%: 0.00002–0.00830) for IFN-α, again indicating the goodness of fit of our analytic approach.

Correlating growth and regression rate constants with OS in patients treated with sunitinib or IFN-α
Previous analyses (9–12) suggested the growth rate constant, g, could be added to response rate and progression-free survival (PFS) as a measure of efficacy. Here, the median g of tumors in patients receiving sunitinib (0.00082 per day; log g = −3.09) was 45% lower than that (P < 0.001) of tumors in patients receiving IFN-α (0.0015 per day; log g = −2.81).

Recognizing OS as the “gold standard” of drug efficacy, we assessed the correlation between g or d and OS. The final analysis of OS (13) was used, with 262 OS events [date of death (DOD) documented] that yielded acceptable g/d parameters (including 226 with acceptable g). Figure 2 top, left shows the OS/log g correlation for the 131 sunitinib patients having both a valid g (P < 0.1) and DOD. The right top panel is a similar plot for the 95 patients treated with IFN-α. For sunitinib, log g and OS were significantly negatively correlated (Rsq = 0.44; P < 0.001), whereas the data for IFN-α (Rsq = 0.14; P < 0.001), were less strongly correlated. These g values were calculated using tumor measurements obtained by study investigators. Regressions using measurements from central review were similar: for sunitinib, log g and OS correlated significantly negatively (Rsq = 0.30; P < 0.001) in contrast to those for patients randomized to IFN-α (Rsq = 0.14; P < 0.0044). The bottom panels of Fig. 2 compare regressions of tumor nadir (the measure used in response rate, defined as the ratio of minimum sum of LDs to initial sum of LDs) and PFS or OS in the sunitinib arm. In prostate cancer (10), nadir depth and time to reach nadir were both surrogates of g, faster growth rates producing higher nadirs, shorter times to nadir, and, in turn, shorter PFS (10). Accordingly, lower correlations (Rsq = 0.19) were found when nadir, rather than log g, is regressed on OS in this data set, the nadir being merely a surrogate of g. The regression of PFS with OS was comparable with the results obtained with g (data not shown).

To test whether patients who died while on study were representative of the whole population, Supplementary Fig. S2 shows log g versus OS for all patients for whom we had valid g parameters as solid circles, with the subset that had died as red open circles. The data sets fall on a continuum, suggesting that deceased patients are indeed representative of the whole population insofar as dependence of survival on g is concerned.

A stepwise regression of OS on log g, log d, and initial tumor quantity (f₀, sum of LDs) showed that only log g contributed to the regression for patients randomized to sunitinib (P < 0.001), whereas both log g (P = 0.010) and f₀ (P = 0.028) contributed significantly to OS in patients randomized to IFN-α (data not shown).

Comparing growth rate constants extracted from the two study arms
The 2 panels on Fig. 3 left, depict waterfall plots of tumor fraction after 12 treatment weeks as percentage of the sum of the LD at enrollment and show that sunitinib is more
effective than IFN-α. Because calculations of $g$ and $d$ use the time elapsed between assessments, comparisons among studies can be made regardless of differences in assessment protocols. Figure 3 right, depicts dot plots of log $g$ values derived in our previous studies in patients with mRCC, treated with a placebo, bevacizumab, or ixabepilone (9, 11, 15, 16) as well as the present 4 data sets (sunitinib or IFN-α, each measured by study investigators or central review). Confirming the waterfall plots, $g$ values for sunitinib-treated patients are significantly slower at $P < 0.0001$ [investigatorscentral review log $g$ values were: sunitinib $= -3.09/\pm -2.94$; IFN-α $= -2.81/\pm -2.78$].

Comparing rate parameters derived from data obtained by study investigators or by independent central review

Figure 4 is a direct comparison of results obtained using data from study investigators or central review. (Log $g$ values calculated using the central review on the $x$ axis, corresponding log $g$ values using study investigator measurements on the $y$ axis). Rsq for the plot is 0.80; the regression slope is $1.14 \pm 0.04$, close to unity. Accordingly, no statistical advantage was gained by including central review data.

g and $d$ parameters can be extracted with accuracy before the nadir is reached

The 8 panels of Fig. 5, left, show successive values of the sum of LDs from 1 patient, obtained over time with starting quantity arbitrarily set at 1. In each plot, one additional time point is added. Applying Eq. (A) to the data in each plot, we obtained intermediary values for $g$ and $d$ (mean ± 95% CI), plotting the values in panels on the right. By the fourth reevaluation both $g$ and $d$ are obtained with accuracy; later values differ little. Thus, $g$ can be accurately estimated long before tumor measurements show growth: the downward trajectory is “deviated upward” by the as yet “undetected” growing fraction. In similar analyses of 20 randomly chosen patients, an accurate estimate of $g$ could, in most cases, be obtained before the data showed a more than 20% rise in the sum of LDs (not shown). For these 20 cases, the time between recording a $g$ estimate and either a more than 20% rise in LD sum above the nadir or the end of assessment, was a median of 126 days (25%,–75%: 84–169 days). Thus, 4 months before treatment is discontinued, the tumor growth rate can be discerned.

The stability of the derived $g$ parameter was examined, carrying out the same procedure on the entire data set. For the overwhelming majority of cases, $g$ remained constant; only rarely was a final $g$ value significantly above previously determined values. The first 8 cases in each arm with valid $g/d$ values, in which data collection had continued above 320 days, are depicted in Supplementary Fig. S3.

Estimating the time to death, had treatment continued beyond the arbitrary definition of progression

Patients (with known DOD) were divided into 5 groups (quintiles), of increasing $g$. The highest quintile comprises patients with no apparent benefit from sunitinib, whose on-study $g$ is likely similar to their off-treatment $g$. Substituting median $g$ and $d$ values into Eq. (A), using DOD as ($t$), we...
estimated the relative "tumor size" at time of death. For the highest quintile, the median relative tumor size (sum of LDs, f) at DOD was 2.1-fold that of the median entry value (sum of LDs) for the whole population (for a sphere this represents a 9.3-fold increase in tumor volume). Assuming groups of patients die with similar tumor burdens (some reaching this faster than others), we used 2.1-fold the entry value as the tumor quantity at death. In Fig. 6, we used median g and d values, in quintiles, for sunitinib-treated patients, to determine the time at which the projected tumor growth curve intersects a final expected tumor quantity of 2.1-fold. For the slowest growing quintile (largest negative log g value), in which most patients showed only regression while on study, the model projects no tumor growth, although median OS was 516 days (We did not depict this quintile because the g values of nearly zero would give a large error in tumor growth projections). In the next 2 quintiles (A and B), the projected OS based on the on-study g and d values is considerably longer than actual OS. Our projected tumor curve assumes a constant g until the 2.1-fold tumor quantity is reached, so the shorter actual OS implies that g increased after discontinuation of drug. We hypothesize that patients with slow g values could have lived closer to the predicted OS had treatment continued beyond the conventional "20% above nadir" threshold. In C, the second fastest growing quintile, predicted OS is 377 days, close to the 322 found, suggesting the on-study g accurately predicts the g when therapy ceases, perhaps indicating drug resistant tumor. For the
Discussion

In this study, using data from the pivotal trial of sunitinib versus IFN-α (3, 13), we show that analysis of data routinely gathered during a clinical trial can provide estimates of the rates of tumor regression (d) and growth (g). The growth rate constant is an excellent "surrogate" for OS, correlating significantly with OS (P < 0.001); d correlates less well. Both g and d can be extracted, and tumor response characterized, even before the nadir is reached. Sunitinib results in a significantly slower rate of tumor growth (P < 0.0001) as compared with IFN-α, the sunitinib rate being almost better (P = 0.053) than that achieved with bevacizumab in a smaller randomized trial (15). The median g of tumors in patients receiving sunitinib (0.00082 per day; log g = −3.09) was about half that (P < 0.001) of tumors in patients receiving IFN-α (0.0015 per day; log g = −2.81). This is consistent with the reported best response rates determined by RECIST, 47% for sunitinib, compared with 12% for IFN-α (3, 13). The correlation of g with OS observed in patients randomized to sunitinib is consistent with the strong correlation of g with OS that has been observed in our previous analyses (9–12; and in several additional histologies currently under study3), but was less so in those randomized to-

3Multiple myeloma, and breast and thyroid cancer (work in progress).

IFN-α. This may be due to responses in patients on the IFN-α arm who went on to receive sunitinib after disease progression. Although more patients had g/d values determined than had PFS scored (319 vs. 204 for sunitinib and 240 vs. 186 for IFN-α), one caveat is worth noting. Patients on the IFN-α arm were more likely to have "unacceptable" g values (63 had only 2 evaluations with <1.2-fold difference) than those on sunitinib (26 had only 2 evaluations with <1.2-fold difference; Supplementary Fig. S1), potentially introducing the same type of bias that can occur in Kaplan–Meier projections when patients are censored before progression.

We envision the g value to be particularly useful as a clinical trial endpoint that could provide an independent analysis of drug efficacy. Using SigmaPlot’s size function in its statistics package, and the SE of 0.0019 per day found for the g values in the sunitinib arm, 100 cases in each of 2 arms would be needed in some future trial to find a decrease of 50% in the mean of g with a power of 0.85 at P = 0.05. We also compared tumor measurements determined by study investigators and by central review and observed that, despite absolute differences in measurements, the rate of change of these measurements—that is, growth rate, g—is nearly identical, and their correlation with OS indistinguishable. These analyses are in agreement with prior reports of overall concordance of independent and investigator review and suggestions that such review does not add value (17, 18). As an efficacy endpoint, g could support investigator assessments and help eliminate need for central review.

Despite the overall correlation of on-study g with OS in the sunitinib arm, we could model a change in growth rate.
after study discontinuation in some patients who had marked slowing of growth while on study. Using the OS of enrolled patients and the derived $g$ and $d$ parameters, we estimated a relative tumor size at death 2.1-fold that at the time of enrollment (9.3-fold increase in tumor volume) in patients who did not benefit from therapy. This value is similar to that determined in a previous mRCC analysis (11.5-fold increase in tumor volume). On the basis of growth rates determined while on study, patients with slower on-study $g$ values should have lived longer. The model suggests tumor growth rates increased after treatment was discontinued, to approximate that of a placebo rate [faster than the rate on therapy, but one that would not be characterized as “acceleration” (19)]. We found no evidence that $g$ accelerated as treatment continued. As shown in Fig. 5 and Supplementary Fig. S3, on-study $g$ did not rise appreciably—indeed at the time of treatment discontinuation, the value of $g$ was comparable with that obtained months earlier.

An important consequence of the survival projection follows: continuation of sunitinib treatment beyond study criteria for progression might have extended survival substantially in patients whose $g$ values indicated substantial slowing of tumor growth. For example, for the 30% of patients with the slowest calculable $g$ value, the model predicts survival to 1,365 days (to reach 2.1-relative size), while the actual median OS and 25% to 75% limit was 633 (523–777) days. Because we lack curative therapies, using the $g$ value to predict who might benefit from continued treatment, provided it is clinically feasible and tolerable, offers the possibility of prolonging survival, in the absence of good salvage or curative therapies—a hypothesis that needs to be prospectively tested.

Several caveats should be noted about the survival analysis. First, the survival modeling was carried out only in patients who had died, rather than the entire study population; once the study was closed, collection of survival data ceased. Thus, a survival benefit due to extended therapy may be only seen in a subgroup with more advanced disease—the measured parameters in Supplementary Table S1 certainly suggest that the patients in the trial population who had died had more advanced disease, although the analysis of $g$ in the 2 groups in Supplementary Fig. S2 suggests dependence of survival on growth rate was similar whether patients were living or had died. Second, because many patients receiving IFN eventually received other therapies including sunitinib and other VEGF inhibitors, this may have reduced the survival difference between the 2 arms. In a similar way, the multiple lines of salvage therapy that are now available for renal carcinoma could confound the interpretation of a study aimed at proving the benefit of extending sunitinib therapy in renal cancer.

Finally, we would observe that in mRCC we have entered the era of “targeted therapies” leaving behind IL-2 and IFN-$\alpha$, without understanding who benefited and why (20). Our analysis concurs with a recent review that concluded “genuine, if modest, effectiveness” of IFN-$\alpha$ in mRCC (21), in that there were a number of patients in the IFN-$\alpha$ cohort...
with very slow growth rates—although proof that this could be ascribed to IFN-α requires pretreatment tumor measurements.

In summary, we show that a tumor’s growth rate constant can be reliably calculated from clinical data gathered during a randomized phase III trial and that this value can reliably predict the FDA gold standard, OS. We confirm the efficacy of sunitinib versus IFN-α and compare it with other agents used in mRCC. A similar analysis of data from other trials could reliably compare all existing therapies for mRCC. As regards individual patients, we show that very early in treatment, long before disease progression is scored, indeed before tumor growth is observed, one can determine a growth rate constant. This knowledge enables a reliable estimate of the predicted OS, which could be used in clinical trial designs to continue therapy past the arbitrary 20% progression in prostate cancer trials: the growth rate as an indicator of therapeutic efficacy. Clin Cancer Res 2011;17:907–14.


Analyzing the Pivotal Trial That Compared Sunitinib and IFN-α in Renal Cell Carcinoma, Using a Method That Assesses Tumor Regression and Growth

Wilfred D. Stein, Julia Wilkerson, Sindy T. Kim, et al.


Updated version
Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-11-2275

Supplementary Material
Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2012/02/17/1078-0432.CCR-11-2275.DC1

Cited articles
This article cites 20 articles, 10 of which you can access for free at: http://clincancerres.aacrjournals.org/content/18/8/2374.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at: /content/18/8/2374.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.