Published in final edited form as:

Clin Pharmacol Ther. 2013 April ; 93(4): 345–351. doi:10.1038/clpt.2012.263.

Estimation of renal cell carcinoma treatment effects from disease progression modeling

Michael L. Maitland^{1,2,3}, Kehua Wu¹, Manish R. Sharma^{1,2}, Yuyan Jin⁵, Soonmo Peter Kang^{1,2}, Walter M. Stadler^{1,3}, Theodore G. Karrison^{3,4}, Mark J. Ratain^{1,2,3}, and Robert R. Bies⁶

¹Department of Medicine, Section of Hematology/Oncology

²Committee on Clinical Pharmacology and Pharmacogenomics

³Comprehensive Cancer Center

⁴Department of Health Studies, University of Chicago

⁵Department of Pharmaceutical Sciences, University of Pittsburgh

⁶Division of Clinical Pharmacology, Indiana University

Abstract

To improve future drug development efficiency in renal cell carcinoma (RCC), a disease progression model was developed with longitudinal tumor size data from a phase III trial of sorafenib in RCC. The best fit model was externally evaluated on 145 placebo-treated patients in a phase III trial of pazopanib, and incorporated baseline tumor size, a linear disease progression component, and an exponential drug effect parameter. With the model-estimated effect of sorafenib on RCC growth we calculated the power of randomized phase II trials between sorafenib and hypothetical comparators over a range of effects. A hypothetical comparator with 80% greater drug effect than sorafenib would have 82% power (one-sided $\alpha = 0.1$) with 50 patients per arm.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Michael L. Maitland 5841 S. Maryland Ave. MC-2115 Department of Medicine Section of Hematology/Oncology Chicago, IL 60637 Telephone: 773-834-8981 Fax: 773-702-9698 mmaitlan@medicine.bsd.uchicago.edu.

Author Contributions

MLM, KW, MRS, YJ, WMS, TGK, MJR, and RRB wrote the manuscript MLM, MRS, SPK, WMS, TGK, MJR, and RRB Designed the Research MLM, KW, MRS, YJ, SPK, TGK, and RRB Performed the Research MLM, KW, MRS, YJ, TGK, MJR, and RRB Analyzed the Data KW, MRS, YJ, TGK, and RRB Contributed Analytical Tools

Y.J. current address Clinical Pharmacology Research Unit, Pfizer, Groton, CT

S.P.K. current address Clinical Oncology Research Unit, Merck, Kenilworth, NJ

MLM and MRS contributed equally to this manuscript.

Preliminary data from this study were presented, in part, at the American Society of Clinical Pharmacology and Therapeutics Annual Meeting, 2011 in Dallas, TX.

Conflict of Interest/Disclosure

MLM has performed compensated consulting for Astellas Pharma U.S., and Pfizer, Inc.. YJ is currently employed by Pfizer, Inc with stock ownership. SPK is currently employed by Merck & Co. in which he has stock ownership. WMS has performed consulting for Novartis, Pfizer, Inc., Roche/Genentech, and Aveo and has been on speakers' bureaus for Bayer, Inc. and Pfizer, Inc. MJR reports a compensated consulting relationship with Mylan. RRB has received support from the Indiana CTSI through a gift from Eli Lilly and Company as well as grants from Merck through the Regenstrief Institute and NICHD.

Model-based quantitation of treatment effect with CT imaging offers a scaffold on which to develop new, more efficient, phase II trial endpoints and analytic strategies for RCC.

Keywords

Antineoplastic Agents; Antiangiogenic Agents; Carcinoma; Renal/drug therapy; Decision Making; Drugs; Investigational; Humans; Models; Statistical; Renal Neoplasms/drug therapy

INTRODUCTION

Novel anticancer therapeutics fail in phase III trials with higher frequency than drugs developed in other fields of medicine (1-2). Because phase III trials are the most expensive to conduct, this high failure rate disproportionately increases the total costs for development of anticancer therapy. The need to modernize decision-making in the phase II to phase III transition has been recognized for nearly a decade(3), and the issue of improving phase II clinical trial design has garnered increasing attention(4-12).

Measuring the effects of treatment for solid tumors has been difficult to standardize(13-15). The most widely used method, the Response Evaluation Criteria in Solid Tumors (RECIST) (16), achieves reproducible results by estimating the quantitative change in tumor burden and then categorizing disease response as: complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). The cut-points to separate disease response into the different categories were intended to minimize miscategorization and were first established when digital imaging was a new technology. At that time there were few available standard treatments for advanced solid tumors. A categorical system was efficient for screening new agents, especially drugs that were expected to shrink tumors within short timeframes. Now that solid tumors have become more treatable, and testing novel agents in combination programs or in subsets of patients has become more common, the inefficiencies of categorical tumor assessment are becoming less sustainable(11, 17).

Efforts to improve phase II trials are especially timely for RCC where 7 new agents in 6 years have been FDA-approved for metastatic disease(18-19). A promising approach to evaluate new therapies, recently demonstrated in lung and colorectal cancer(20-22), is to model the course of disease progression from tumor measurements collected in previous clinical trials. Relying upon the sum of the longest dimensions of the measured target lesions rather than RECIST categorization, this approach offers an efficient, adaptable method for evaluating treatments with the current commonly used, primarily CT-imaging data.

As each disease has different, characteristic, patterns of growth, we have developed a longitudinal tumor growth model for RCC based on measurements collected in the phase III trial that supported regulatory agency approval for sorafenib (23). We then validated this model with measurements collected in the phase III trial that led to approval of pazopanib(24). To determine the potential for evaluating therapeutics based on a quantitative assessment of treatment effect we performed simulations to assess the power of randomized phase II trials to detect different magnitudes of treatment effect above the estimated effect of sorafenib.

RESULTS

Longitudinal growth model of RCC

In the TARGET study sample used to generate the initial model, 36% of patients had one tumor lesion, 20% had two, and 44% had the sum of three lesions available for serial measurement. Initially, linear and exponential models were used to fit the placebo data for estimating tumor growth rate. The linear model described the data better than the two exponential models (Table 1), with lower value of Akaike information criterion (AIC) and lower objective function value (OFV). Models with a drug effect (DE) parameter were explored with the placebo data but no such parameter could be estimated (lack of convergence), suggesting that there was no average reduction in tumor size in the placebo arm.

The final model simultaneously estimated mean baseline tumor size (BASE), the tumor growth rate parameter (PR), drug effect (DE) in the sorafenib arm and residual variance using all available data (both placebo and treatment arms) of the TARGET study. To estimate sorafenib's effect on renal cell carcinoma growth entailed an exponential treatment effect parameter. Estimated parameters and their relative standard errors (%SEs) are presented in Table 2. Note that NONMEM fits a non-linear, mixed effects model incorporating the fixed effects (BASE, PR, DE) as well as random, inter-individual variability in these parameters. The %SEs reported from NONMEM are all less than 25%.

One artifact of hierarchical modeling approaches such as this one is known as Bayesian “shrinkage.”(25) Excessive shrinkage could lead to spurious identification of patient-specific factors associated with tumor progression. Although we did not pursue detailed evaluation of patient-specific factors, we calculated shrinkage in interindividual variability (η), which is reported in Table 2. As clinical and molecular differences among patients and their tumors respectively might lead to distinct observable differences in baseline tumor size and progression rate, mixture distributions were explored for BASE and PR parameters in the placebo dataset. However, these models were not stable and model diagnostics reflected significant mis-specification with the mixture distribution models. Thus our model makes the standard assumption that the random effects follow a normal distribution. Finally, bootstrapping was also used to evaluate the uncertainty in parameter estimates, with median values from 1,000 replicates shown (Table 2). The median parameter values resulting from the bootstrap procedure agreed with the estimates from our final model. This suggests that the parameters in the final model were reasonably well determined and the model was stable. From 2000 bootstrap runs, 1000 (placebo model) and 947 (combined placebo and sorafenib model) minimized successfully and were included in the bootstrap analysis.

Consideration of informative dropout

In these studies, investigators used their clinical assessments and RECIST criteria to determine when patients had disease progression. This means that the patients' tumor measurement data could be censored without additional CT measurements (clinically determined progression), or with solely new, small, but difficult-to-measure lesions that indicate RECIST progression, but not contribute significantly to the sum of the longest

dimensions of tumor measurements. The probability of dropout (termination of the longitudinal data due to disease progression or censoring) is therefore dependent on this unobserved tumor growth since the most recent RECIST measurements. To determine whether this non-random dropout affected the model we estimated the dropout hazard (β_2), which indicated an increased hazard of dropping out with larger tumor size. The last observation was also considered to be related to withdrawal, but the dropout model did not support inclusion of both the last observation and the unobserved tumor size due to their high correlation. With lower objective function value and shrinkage, the unobserved predicted tumor size was included in the dropout model. However, this dropout model had minimal effect on the renal cancer growth model parameter estimates (Table 2). This suggests that the development of new metastatic lesions does not contribute to significant deviation from predictions of this linear growth model. The likelihood of disease progression is captured by the increasing sum of longest dimensions of the original target lesions.

Model performance and external validation

Selected cases of placebo-treated and sorafenib-treated patients demonstrate agreement between predicted and observed data (Figure 1). We performed an external validation of the model with data collected from 145 patients assigned to placebo in the randomized phase III trial of pazopanib. The visual predictive check (VPC) (Figure 2) depicts the distribution of the sum of the largest dimensions of all target lesions for each patient in the placebo arm over time. The median, 5th, and 95th percentiles of the tumor size distributions are captured by the model-based predictions through approximately 168 days (or 4 CT imaging evaluations). Importantly, the model captures the population “dip” in tumor size at the 95th percentile over the first, second, and third CT imaging evaluation intervals as the subjects with the largest tumor burdens drop-out.

Power calculations for drug effect (DE) parameter as an endpoint in future trials

The DE parameter in the tumor growth model provides a quantitative representation of the effect of drug on tumor growth on a continuous scale. By definition, the placebo takes a value of $DE = 0$. In a placebo-controlled trial the difference in drug effect = $DE_{\text{treatment}} - 0$. For sorafenib in the TARGET trial, $DE = 0.00443/\text{day}$. One way to interpret the model is that for a patient with a baseline tumor size of 62.7 mm in the placebo arm, on average the tumor will increase to $62.7 + 0.158 * 30 = 67.4$ mm after one month, a 7.5% increase. In the sorafenib arm, after one month the expected size is $62.7e^{-0.00443 * 30} + 0.158 * 30 = 59.6$ mm, a 5% decrease. This may not be as intuitive to the trialist as progression-free survival, but is a novel quantitative term to reflect the impact of drug in terms of % reduction in tumor size. In the TARGET trial, HR for disease progression was 0.44 (5.5 months sorafenib arm, vs. 2.8 months in the placebo arm). A larger DE will result in a lower HR.

As standard-of-care treatments become available and serve as comparators, clinical trials can be designed to detect evidence of improvement in DE over that achieved with sorafenib. Given the variance on the DE estimate for sorafenib from the 749 subjects for whom data were available, we assessed power for relative improvements in the drug effect over sorafenib in simulated, randomized, 2-arm, phase II trials having 50 patients per arm with a

one-sided Type I error rate (α) = 0.1 (Figure 3). The ratio of DE for the comparator compared to sorafenib ranged from 1.2 to 2.0 for which power ranged from 30-91%. Randomized studies (50 patients/arm) had 82% power to detect a significant difference in drug effect when the comparator had a true DE 1.8 times (or 80% greater than) that of sorafenib, and 91% power to detect a significant difference in DE when the comparator had a true DE 2.0 times (or 100% greater than) that of sorafenib.

DISCUSSION

We have presented a modeling framework to quantify the magnitude of DE for therapeutics tested in advanced RCC, based on CT imaging measurements collected in phase III trials. The model was first developed with imaging data from the multicenter trial of sorafenib in RCC, and then validated with imaging data from the multicenter trial of pazopanib in RCC. Simulations using the DE model parameter suggest that a randomized 2-arm trial (with 50 patients/arm) would have sufficient power to detect new treatments with at least 80% greater DE than sorafenib. As with any endpoint, larger clinical trials would be required to detect smaller improvements in DE over sorafenib.

This initial modeling effort to approach RCC progression evaluation on a continuous rather than categorical scale takes into account the full set of longitudinal measurements. This model serves as a modifiable scaffold upon which new advances in predictive and prognostic factors could be incorporated to make clinical trials in advanced RCC more efficient. The model that best fit this large dataset relies on three fundamental parameters: the baseline burden of disease represented by the sum of the longest dimensions of the measurable lesions on CT imaging (BASE), and the combination of DE and PR, reflecting the growth rate determined by the changes in total tumor burden assessed over time. It is intuitively obvious that differences in “how much tumor you have” and “how fast it is growing” are the primary variables for assessing treatments in advanced disease. Indeed these are the same parameters used in disease progression models for colorectal cancer and non-small cell lung cancer (20-21). Based on readily collected empirical imaging data, this model doesn't address the more theoretical, mechanistic considerations of tumor fractions growing and regressing during the course of treatment(26). However, models of this structure have been robust for application to both first-line and second-line treatment settings in other tumor types (27). The BASE parameter is typically larger in second-line than first-line treatment, but the linear progression rate is within the same range. As these empirical models have recently been developed, potentially important predictors such as prior therapy response, molecular markers, and co-morbid conditions have not yet been incorporated.

For the model in this paper, we referred to the exponential decay parameter as drug effect (DE). In the non-small cell lung cancer model, developed primarily with cytotoxic therapeutics, this term is called the shrinkage rate (SR)(20), and in the colorectal cancer model, developed solely with cytotoxic therapeutics, the term is drug constant cell kill rate (K_D)(21). This RCC model has been developed and validated solely with inhibitors of VEGFR2. These agents may decrease the measured size of tumors in part through killing tumor cells, but slowing the pace of further tumor growth, and diminishing intratumoral

hydrostatic pressure are other means by which these drugs may decrease size and rate of progression of tumor masses (which can reduce associated symptoms). This therefore justifies the more general term, DE.

As for many different solid tumor types, the advancement of therapy for advanced RCC has entered a new era. Several distinct chemical entities have clear evidence of benefit compared with placebo(28), and more recently these have begun to be compared directly against established standard therapies (19). However, the need to detect evidence of clinically significant benefit for new drugs in disease subsets, and to detect benefit of combination treatments with factorially increasing combination possibilities requires new approaches to the conduct of phase II clinical trials(7, 17, 29). For RCC the disease progression model in this study may improve clinical research efficiency in a number of ways. First, as in the conduct of trials using RECIST, the solely required technology is readily available CT imaging. Second, studies to detect meaningful differences in treatment effects among randomly assigned groups are likely to require the fewest patients with a quantitative parameter endpoint for which inter-subject variance can be estimated and incorporated. Third, this and subsequent models can be shared and optimized publicly. For example, if a tumor tissue marker (such as *VHL* genotype) or novel imaging modality (such as PET or CT volume measurements) significantly improves the capacity to assess tumor growth response to treatment, that parameter could be readily incorporated into the model and used by investigators in subsequent trials.

The findings from this modeling study are consistent with the results from other efforts and support innovative approaches to new endpoints in RCC clinical trials. Simulations of data from completed phase III clinical trials of pazopanib(30) and sunitinib(31) have suggested improved power in phase II trials of RCC for alternative endpoints to standard response rate or PFS based on RECIST. These studies and our investigation, all conducted on VEGF-signaling pathway inhibitors, support the use of a change in tumor size measurement at a fixed early timepoint, such as 8 weeks as a more sensitive endpoint for testing new RCC therapeutics than the conventional RECIST response rate. By obviating the need for the longer term follow up of PFS, studies with this endpoint might also be completed more rapidly and inexpensively. Although a model-based endpoint of DE promises additional efficiency it would be simplistic and premature to compare its performance characteristics with PFS. As recently published in this journal(32), a complete modeling and simulation study, incorporating data from multiple drugs tested in a randomized trial would be needed to assess the relationship between the model parameter, DE, and PFS.

There is room for improvement upon this initial model-building effort. Due to the limitations of the studies in which imaging data were collected, the model describes tumor growth and not patient survival patterns. However, Heng, et al.(33) have suggested that progression-free survival and overall survival are well correlated. The variances not accounted for by the model include not only disease heterogeneity, but also the flaws in the data collection and transmission process. Although the model structure would be unlikely to change significantly, the parameter estimates would likely be more precise if measurements collected directly from images by a central reviewer were used rather than data extracted from radiology reports and transcribed by research associates. Finally, a recurrent criticism

of models derived from sum of longest dimensions measures is that disease progression defined by new lesions is inadequately captured. Our assessment of an informative drop-out effect suggests this might be relevant and worth further development but does not impact the power of this model to detect meaningful treatment effects in a modest-sized cohort of patients randomly assigned to different treatment arms.

Although solid tumor progression models may share the same general structure across disease types, the estimated parameters of these equations will be disease, and disease sub-type specific. Analyzing data in two of the largest advanced RCC trials ever conducted, this model implies that further advances in development of RCC therapeutics may be made through serial studies that collect CT imaging measurement data to examine the quantitative estimates of drug effect. This model enables further refinements and improvements through testing of novel variables and contributions of novel technology. By relying upon routinely collected imaging data (CT scans), this model offers a novel opportunity to advance clinical research in RCC.

METHODS

Trials and Data

Model development was performed with tumor measurement data, as collected and reported by investigators from 749 patients (375 placebo arm; 374 sorafenib arm) enrolled in the phase III Treatment Approaches in Renal cancer Global Evaluation Trial (TARGET)(23). According to local standards for adherence to RECIST(15), target lesions were identified and measured by investigators. To represent the typical quality of smaller scale, locally performed, phase II trials, the complete, independently reviewed imaging data were not used. Model validation was performed with centrally reviewed data from the 145 patients assigned to placebo treatment in VEG105192, the multicenter phase III study of pazopanib(24). In both studies, imaging assessments were conducted every 6 weeks for the first 24 weeks and every 8 weeks thereafter. Conduct of these studies was approved by the University of Chicago Institutional Review Board before the project commenced.

Data extraction

To maximize informativeness of investigator-level data, we generated a program script in the R software environment (Supplemental Information) to extract tumor measurements in a format most consistent with RECIST 1.1(16). The script produced a dataset for import into the nonlinear mixed effects modeling software NONMEM (version VII, level 1, ICON, Ellicott City, MD, USA)(34). This dataset headings included: patient ID, visit, lesion #, lesion size, sum of lesion sizes by RECIST 1.1, sum of lesions, a flag for whether or not a lymph node measurement was incorporated, and the number of the visit. New lesions were identified, but not incorporated into the initial model building. Patients were also flagged for clinical progression without a confirmatory CT scan. If a lesion decreased to 0 and was not tracked consistently, that lesion was not included in the patient's sum of longest diameters, unless it was the only baseline lesion.

Tumor growth and sorafenib effect model development

The longitudinal tumor size data (sum of the longest diameters of the serially tracked target lesions) were fitted with nonlinear mixed effects modeling. The model was built in three stages. Initially, both linear and exponential models were used to fit the placebo dataset for estimating tumor growth rate. The second step was to add the drug effect data and estimate the drug effect parameters with the placebo effects fixed. The final model simultaneously estimated baseline tumor size (BASE), the tumor growth rate parameter (PR), drug effect (DE) and residual variance on all available data (both placebo and treatment arms) of the TARGET study.

The first-order conditional estimation (FOCE) algorithm for calculating the likelihood in NONMEM was used for parameter estimation. The tumor progression model describes tumor size as a function of time and accounts for the natural growth of the tumor and drug effect (DE). Linear, exponential and power functions were used to fit tumor size data as candidate models describing both tumor shrinkage and tumor growth. Mixture distributions were explored to model potential subgroups with different tumor growth and/or responses to treatment. A model using a combination of an exponential-decay (shrinkage) and linear-growth (progression) was developed, as it was the best for describing the tumor growth and drug effect. Model structural selection was guided using the objective function value difference as well as the AIC. An alpha value for significance was set a-priori to $\alpha < 0.01$ (objective function value difference of 6.63 for one degree of freedom). Akaike Information Criterion was also used to adjust for degrees of freedom but the absolute difference was assessed and not statistical significance.

The final model fitted is

$$TS_i(t) = BASE_i \cdot e^{-DE_i \cdot t} + PR_i \cdot t \quad (\text{Eq. 1})$$

$TS_i(t)$ is the tumor size at time t for the i th individual, $BASE_i$ is the baseline tumor size, DE_i is the exponential tumor shrinkage rate parameter due to the drug effect and PR_i is the tumor growth rate constant, all for the i th individual. $BASE_i$, PR_i , and DE_i incorporate random deviations for the i th individual about the respective population mean parameter. For additional details of model development see Supplemental Methods.

The final tumor growth model parameter estimates were further evaluated internally using a nonparametric bootstrap. The resampling was performed 1000 times. The median values and the 2.5th and 97.5th percentiles of the parameter estimates obtained by this analysis were compared with those of the final model.

Model Validation and Power Assessment

External model validation—The model was evaluated with visual predictive check (VPC)(35-37) of the 145 placebo-treated subjects from the VEG105192 multicenter phase III study of pazopanib as described above(24). These placebo data were not included in the initial model development. The VPC was generated using 1000 simulations from the joint tumor growth and dropout model to assess the predictive performance. A graphical

comparison was made between observed data and the model predicted median and 90% prediction interval (90% PI).

Power calculations with drug effect as the endpoint—Randomized, two-arm (50 patients per arm) phase II trials comparing sorafenib and a hypothetical comparator (with drug effect as the primary endpoint) were simulated (with 1,000 replicates) to estimate the power to detect a significant difference between arms ($\alpha = 0.10$). Specifically, simulated data for tumor size at 6 weeks, 12 weeks, 18 weeks and 24 weeks were generated using the baseline tumor size and progression rate from the validated placebo model, with drug effects ranging from 0% to 100% greater than that for sorafenib [0.00443 (sorafenib effect), 0.005316, 0.006202, 0.007088, 0.007974, 0.00886 (twice the sorafenib effect)]. Simulated data used the same estimates of interindividual variability and residual error as fitted for sorafenib. Population estimates of drug effect for the two 50 patient arms in each simulated trial (hypothetical comparator vs. sorafenib) were compared using a z-test, and estimated power was the percentage of trials with a statistically significant difference between the two arms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Sibyl Anderson, Frank Cihon, Chetan Lathia and Gloria Hofilena at Bayer, and Bernard Escudier for sharing the TARGET sorafenib data. James Jiang provided initial assistance with evaluation of investigator-level tumor measurement data. We also appreciate Ohad Amit and Lini Pandite at Glaxo SmithKline, and Cora Sternberg for sharing the pazopanib trial data. Ben Suttle and Peter Bonate were generous in sharing their unpublished findings and manuscript.

MLM was supported by National Cancer Institute Mentored Career Development Award K23CA124802. MRS and SPK were supported by T32GM007019 Training in Clinical Therapeutics. This research was also supported by the University of Chicago Cancer Research Center P30-CA014599 and an ASCO Translational Research Professorship (MJR). RRB is supported through the Indiana CTSI (UL1RR025761-01) through a gift of Eli Lilly and Company. Bayer, Inc. and Glaxo SmithKline, Inc. provided tumor measurement data.

REFERENCES

1. DiMasi JA, Grabowski HG. Economics of new oncology drug development. *J Clin Oncol*. Jan 10; 2007 25(2):209–16. [PubMed: 17210942]
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*. Aug; 2004 3(8):711–5. [PubMed: 15286737]
3. Roberts TG Jr, Lynch TJ Jr, Chabner BA. The phase III trial in the era of targeted therapy: unraveling the "go or no go" decision. *J Clin Oncol*. Oct 1; 2003 21(19):3683–95. [PubMed: 14512401]
4. Adjei AA, Christian M, Ivy P. Novel designs and end points for phase II clinical trials. *Clin Cancer Res*. Mar 15; 2009 15(6):1866–72. [PubMed: 19276272]
5. Brown SR, Gregory WM, Twelves CJ, Buyse M, Collinson F, Parmar M, et al. Designing phase II trials in cancer: a systematic review and guidance. *British journal of cancer*. Jul 12; 2011 105(2): 194–9. [PubMed: 21712822]
6. Gan HK, Grothey A, Pond GR, Moore MJ, Siu LL, Sargent D. Randomized phase II trials: inevitable or inadvisable? *J Clin Oncol*. May 20; 2010 28(15):2641–7. [PubMed: 20406933]

7. Mandrekar SJ, Sargent DJ. Randomized phase II trials: time for a new era in clinical trial design. *J Thorac Oncol*. Jul; 2010 5(7):932–4. [PubMed: 20581575]
8. Rubinstein L, Crowley J, Ivy P, LeBlanc M, Sargent D. Randomized Phase II Designs. *Clinical Cancer Research*. Mar 15; 2009 15(6):1883–90. 2009. [PubMed: 19276275]
9. Rubinstein L, Leblanc M, Smith MA. More randomization in phase II trials: necessary but not sufficient. *Journal of the National Cancer Institute*. Jul 20; 2011 103(14):1075–7. [PubMed: 21709273]
10. Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clin Cancer Res*. Mar 15; 2010 16(6):1764–9. [PubMed: 20215557]
11. Sharma, MR.; Maitland, ML.; Ratain, MJ. *Cancer journal*. Vol. 15. Sudbury, Mass.: Sep-Oct. 2009 Other paradigms: better treatments are identified by better trials: the value of randomized phase II studies; p. 426-30.
12. Sharma MR, Stadler WM, Ratain MJ. Randomized phase II trials: a long-term investment with promising returns. *Journal of the National Cancer Institute*. Jul 20; 2011 103(14):1093–100. [PubMed: 21709274]
13. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer*. Jul; 1976 38(1):388–94. [PubMed: 947531]
14. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. Jan 1; 1981 47(1):207–14. [PubMed: 7459811]
15. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute*. Feb 2; 2000 92(3):205–16. [PubMed: 10655437]
16. 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*. Jan; 2009 45(2):228–47. [PubMed: 19097774]
17. Maitland ML, Hudoba C, Snider KL, Ratain MJ. Analysis of the yield of phase II combination therapy trials in medical oncology. *Clin Cancer Res*. Nov 1; 2010 16(21):5296–302. [PubMed: 20837695]
18. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. *Lancet*. Mar 28; 2009 373(9669):1119–32. [PubMed: 19269025]
19. Rini BI, Escudier B, Tomczak P, Kaprin A, Szczylik C, Hutson TE, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet*. Dec 3; 2011 378(9807):1931–9. [PubMed: 22056247]
20. 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. *Clinical pharmacology and therapeutics*. Aug; 2009 86(2): 167–74. [PubMed: 19440187]
21. 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*. Sep 1; 2009 27(25):4103–8. [PubMed: 19636014]
22. Tham LS, Wang L, Soo RA, Lee SC, Lee HS, Yong WP, et al. A pharmacodynamic model for the time course of tumor shrinkage by gemcitabine + carboplatin in non-small cell lung cancer patients. *Clin Cancer Res*. Jul 1; 2008 14(13):4213–8. [PubMed: 18594002]
23. Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *The New England journal of medicine*. Jan 11; 2007 356(2):125–34. [PubMed: 17215530]
24. Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol*. Feb 20; 2010 28(6):1061–8. [PubMed: 20100962]

25. Louis TA, Shen W. Innovations in bayes and empirical bayes methods: estimating parameters, populations and ranks. *Stat Med*. Sep 15-30; 1999 18(17-18):2493–505. [PubMed: 10474155]
26. 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*. Oct; 2008 13(10):1055–62. [PubMed: 18827177]
27. Houk B. Predictive capability of a Food and Drug Administration (FDA) overall survival model in non-small cell lung cancer in two phase II trials utilizing different anti-cancer agents. *Clinical pharmacology and therapeutics*. Feb.2009 (Suppl 1):85. 2009.
28. Vogelzang NJ. Treatment options in metastatic renal carcinoma: an embarrassment of riches. *J Clin Oncol*. Jan 1; 2006 24(1):1–3. [PubMed: 16330665]
29. Maitland ML, Schilsky RL. Clinical trials in the era of personalized oncology. *CA Cancer J Clin*. Nov-Dec;2011 61(6):365–81. [PubMed: 22034206]
30. 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*. Apr 15; 2012 18(8):2309–15. [PubMed: 22287601]
31. 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*. Apr 15; 2012 18(8):2374–81. [PubMed: 22344231]
32. Claret L, Lu JF, Bruno R, Hsu CP, Hei YJ, Sun YN. Simulations Using a Drug-Disease Modeling Framework and Phase II Data Predict Phase III Survival Outcome in First-Line Non-Small-Cell Lung Cancer. *Clinical pharmacology and therapeutics*. Nov; 2012 92(5):631–4. [PubMed: 22910440]
33. Heng DY, Xie W, Bjarnason GA, Vaishampayan U, Tan MH, Knox J, et al. Progression-free survival as a predictor of overall survival in metastatic renal cell carcinoma treated with contemporary targeted therapy. *Cancer*. Jun 15; 2011 117(12):2637–42. [PubMed: 21656741]
34. Beal, SL.; Sheiner, LB.; Boeckmann, AJ. NONMEM Users Guides. Icon Development Solutions; Ellicott City, MD: 2008.
35. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *The AAPS journal*. Jun; 2011 13(2):143–51. [PubMed: 21302010]
36. Post TM, Freijer JI, Ploeger BA, Danhof M. Extensions to the visual predictive check to facilitate model performance evaluation. *Journal of pharmacokinetics and pharmacodynamics*. Apr; 2008 35(2):185–202. [PubMed: 18197467]
37. Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *Journal of pharmacokinetics and pharmacodynamics*. Apr; 2001 28(2): 171–92. [PubMed: 11381569]

Study Highlights

What is the current knowledge on the topic?

Clinical studies to advance cancer therapeutics depend on objective assessment of treatment effects on the pace of growth of tumors.

What question this study addressed?

Longitudinal disease progression modeling offers a quantitative approach to measure tumor burden over time and offers opportunities to detect evidence of biomarker and treatment effects more quickly with fewer patients than current categorical methods.

What this study adds to our knowledge?

We have established and validated a new mathematical model of renal cancer progression based on routinely collected data from two phase III clinical trials in this disease.

How this might change clinical pharmacology and therapeutics?

This model serves as a basic scaffold which can be improved over time and we have demonstrated how the model could be implemented in the design and analysis of a prospective randomized phase II trial.

Fig 1.

Plots of representative individual patients. Placebo-treated patients are in panel A and sorafenib-treated patients in panel B. Circles are observed tumor sizes for the individual patient, solid lines are individual model-predicted tumor sizes, and dashed lines are model-predicted tumor sizes for the entire cohort of placebo- (A) or sorafenib-treated (B) patients.”

Fig 2.

Visual predictive check of the joint tumor size/informative dropout model over 180 days. The circles represent the observed sum of largest dimensions of target lesions for each placebo-treated patient from the pazopanib trial at each timepoint. The solid line is the median of the simulated data, and the 90 percent prediction intervals are encompassed by the dashed lines.

Fig. 3.

Simulation-based power calculation for a randomized phase II trial of a hypothetical new comparator (could be new agent or new agent added to sorafenib) versus sorafenib, using model-estimated drug effect (based on CT scans conducted every 6 weeks up to 24 weeks) as the primary endpoint. We considered a hypothetical new drug with drug effect 1.2 to 2.0 times greater (i.e., 20-100% greater) than that of sorafenib.

Model	AIC	OFV
Placebo model		
TS(t) = BASE + PR × t	7872.4	7868.4
TS(t) = BASE × e ^(PR × t)	7996.8	7992.8
TS(t) = BASE × t ^{PR}	8256.2	8252.2
Combined placebo and treatment model		
TS(t) = BASE – DE × t + PR × t	18459.4	18453.4
TS(t) = BASE × e^{–DE × t} + PR × t	18209.5	18203.5
TS(t) = BASE × t ^{–DE} + PR × t	18534.5	18528.5

Parameter	NONMEM		Bootstrap	
	Estimate	%SE	Estimate	%SE
Tumor progress model				
PR (mm/day)	0.158	11.6	0.158	15.2
BASE (mm)	62.7	2.81	62.7	2.73
DE (1/day)	0.00443	11.1	0.00443	13.9
Inter-individual variability	[shrink %]			
PR (additive)	0.22 [17.9]	21.4	0.23	27.5
BASE (%)	71.5 [3.25]	5.0	71.4	4.80
DE (additive)	0.005 [46.1]	23.2	0.005	27.5
Residual error	[shrink %]			
Proportional (%)	8.9 [30.8]	11.7	8.9	20.23
Dropout model				
Baseline hazard	0.00806	14.6	-	-
Dropout hazard	0.00635	21.6	-	-