



Published in final edited form as:

Cancer. 2014 January 1; 120(1): . doi:10.1002/cncr.28221.

## Progression-Free Survival (PFS) as a Surrogate Endpoint of Overall Survival (OS) in Patients with Metastatic Renal Cell Carcinoma (mRCC)

Susan Halabi, Ph.D.<sup>1,2</sup>, Brian Rini, M.D.<sup>3</sup>, Bernard Escudier, M.D.<sup>4</sup>, Walter M. Stadler, M.D.<sup>5</sup>, and Eric J. Small, M.D.<sup>6</sup>

<sup>1</sup>Department of Biostatistics and Bioinformatics, Duke University, Durham, NC

<sup>2</sup>Alliance Statistical and Data Center, Duke University, Durham, NC

<sup>3</sup>Cleveland Clinic Taussig Cancer Institute, Cleveland, OH

<sup>4</sup>The Institut Gustave Roussy, Villejuif; Centre Léon Bérard, Lyon

<sup>5</sup>University of Chicago, Chicago, IL

<sup>6</sup>University of California, San Francisco, CA

### Abstract

**Purpose**—To investigate the dependence between PFS and OS in mRCC patients and to explore whether PFS can be used as an intermediate endpoint of OS in this patient population.

**Patients and Methods**—A total of 1,381 patients from two prospective phase III trials (CALGB 90206 and AVOREN) of interferon with or without bevacizumab were used in this analysis. Both trials recruited previously-untreated clear-cell mRCC patients with an ECOG performance status 0–2, adequate bone marrow, hepatic, cardiac and renal function and controlled blood pressure. The CALGB study served as the training dataset, and the AVOREN study as the testing dataset. The dependence between PFS and OS was investigated using Kendall’s tau for bivariate time-to-event endpoints.

**Results**—In the training dataset, the median OS among patients who experienced progressive disease at 3 or 6 months were 6 and 8 months, respectively, compared to 25 and 30 (p-value <0.001) months in patients who did not progress. The adjusted hazard ratios were 2.6 (p-value <0.001) and 2.8 (p-value <0.001) for patients who did and did not progress at 3- or 6-months. The dependence between PFS and OS was 0.53. These associations were confirmed in the testing dataset.

**Conclusion**—In mRCC patients treated with interferon with or without bevacizumab, PFS at 3- and 6-months predicts OS. A high dependence between PFS and OS was observed suggesting that PFS may be used as a surrogate endpoint for OS. This is a novel observation for RCC; however these findings require validation in mRCC patients treated with other targeted-agents.

---

Address correspondence to: Susan Halabi, Ph.D., Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2424 Erwin Road, Durham, NC 27705, susan.halabi@duke.edu.

Conflict of interest: Halabi, research funding Pfizer; Dr. Rini: Research and Consulting: GSK, AVEO and Pfizer; Dr. Escudier: Consulting- Bayer, Roche, AVEO, Pfizer, Genentech and Novartis; Dr. Stadler no COI; Dr. Small, none.

Presented in abstract format at the American Society of Clinical Oncology (ASCO) Genitourinary Symposium, San Francisco, CA, March 5-7, 2010 and at the ASCO Annual Meeting, Chicago, IL June 4-7, 2010.

## Introduction

Advances in understanding the genetics and biology of renal cell carcinoma have led to the successful development of novel therapies that target the vascular endothelial growth factor (VEGF) and mammalian target of rapamycin (mTOR) pathways.<sup>1-7</sup> Regulatory approval for almost all of the available agents were based on the rather large improvements in progression free survival (PFS) over a standard control arm. Importantly, in these trials the hazard ratios for PFS for these agents were large, the available standard therapies were sufficiently limited, the toxicity profiles were acceptable in the context of available therapies, and the overall survival (OS) trends mirrored the PFS results in direction even though the magnitude of survival benefit was limited and often not statistically significant. It was thus relatively straight-forward for regulatory authorities to determine that improvements in PFS were reasonably likely to provide a clinically relevant benefit in the context of known risk and known available therapies.

It is critical to note that PFS, although a convenient and objective intermediate endpoint, has not been established as a reliable surrogate endpoint in metastatic renal cancer, even in the context of VEGF pathway directed therapy. This is best illustrated by the recent lack of support for regulatory approval of the alternative VEGF receptor tyrosine kinase inhibitor tivozanib following a phase III trial in which a statistically significant but clinically modest improvement in PFS was demonstrated, but OS favored the control therapy.

These results have re-ignited the debate as to the true relationship between PFS and OS in renal cancer, and whether the observed lack of a tight relationship in phase 3 trials is due to the inadequacy of this intermediate endpoint or the effects of post-trial therapy. The latter is especially important in renal cancer given that several agents were developed and marketed simultaneously.<sup>1-7</sup>

While the relationship between PFS and OS has been evaluated in other tumor types, most notably in colorectal and breast cancer, it has been examined only in one analysis of metastatic renal cell carcinoma(mRCC) patients.<sup>8-11</sup> Heng and colleagues evaluated the relationship between PFS and OS on 1,158 patients treated from 2005–2009. The data, however, were based on retrospective and non-randomized patients and they did not externally validate their findings.<sup>11</sup>

The association of PFS with OS is not a foregone conclusion in mRCC because the development and approval of a large number of agents targeting similar pathways have created the potential for a significant number of post-progression therapies that may confound and attenuate treatment effects on overall survival in phase III trials. Furthermore, the relatively long survival of mRCC patients, and the significant heterogeneity in outcomes, with survival exceeding 10 years in 20% of good prognosis patients, has resulted in the requirement of large trial sizes, long follow-up times and large effect sizes if OS is used as the study endpoint.

It was with these concerns in mind that an evaluation was undertaken of the utility of PFS as an intermediate endpoint for OS in patients with mRCC treated on two large phase III trials.

The main objective of this analysis was to examine whether PFS is a valid intermediate endpoint of OS in patients with mRCC, using data from two large prospective phase III trials (CALGB 90206 and AVOREN) that tested the VEGF targeted agent, bevacizumab, in similar patient populations.<sup>4,5</sup> This allowed the use of one of these studies as a training data set and the second as a testing data set.

## Materials and Methods

### Study Population

Data from CALGB 90206 were used as a training set for this analysis. Between October 2003 and July 2005, 732 patients were randomized to BV 10 mg/kg intravenously every 2 weeks plus IFNA 9 million units subcutaneously three times weekly, or the same dose and schedule of IFNA alone. Randomization was stratified by nephrectomy status (yes, no) and number of adverse prognostic factors (0, 1–2, 3 or more) using established risk criteria.<sup>12</sup> This trial recruited previously-untreated metastatic clear cell RCC patients with an ECOG performance status of 0–2, adequate bone marrow, hepatic, cardiac and renal function, and blood pressure controlled to < 160/90 mmHg. Other eligibility criteria included lack of central nervous system metastases, significant cardiac comorbidity or recent history of bleeding or clotting. The primary endpoint was overall survival with investigator assessed PFS as a secondary endpoint.<sup>4,6</sup>

Data from a similar international phase III trial (AVOREN) were used to confirm the results that were observed in CALGB 90206 and served as the testing dataset. Between June 2004 and October 2005, 649 untreated mRCC patients were randomized. The AVOREN trial had similar eligibility criteria and treatment regimens as CALGB 90206 except that the AVOREN trial used IFNA/placebo for the control group therapy and randomization was stratified by country and number of adverse prognostic factors (0, 1–2, 3 or more). The primary endpoint was investigator-assessed PFS. Details regarding the two trials have been published elsewhere.<sup>4–7</sup>

### Endpoints

The primary endpoint of the present analysis was OS, which was defined as the time from randomization to date of death of any cause. PFS was defined as the time from randomization to date of disease progression or death, whichever occurred first. PFS rate at 3- and 6-months were defined as binary variables: a patient experiencing progression at or before 3-months (or 6-months) was considered as an event. Otherwise, the patient was censored. Radiologic assessments for progression were determined by investigator assessment according to the RECIST criteria at baseline and every 8 weeks on the CALGB protocol. For the AVOREN trial, radiological assessments of progression were performed every 8 weeks up to week 32 and then 12 weeks thereafter until disease progression. Progression was determined by both investigator assessment of radiographs using the RECIST criteria and independent radiological review.

### Statistical Methods and Data Analysis

A similar methodological approach was utilized as reported by Halabi et al.<sup>8</sup> Landmark analyses of PFS at 3- and 6- months from randomization were performed to minimize lead time bias.<sup>13</sup> In the training and testing datasets, 53 and 41 and 133 and 109 patients who died before 3-months and 6-months, respectively, were excluded from the landmark analyses. The Kaplan-Meier product-limit method was used to estimate the OS distribution by the PFS rate at 3- and at 6-months.<sup>14</sup> The proportional hazards model was used to assess the significance of the effect of PFS rate at 3- and at 6- months in predicting OS adjusting for the stratification factors (prior nephrectomy (CALGB 90206), number of adverse risk factors assessed using the MSKCC criteria (CALGB 90206, AVOREN) and country (AVOREN trial).<sup>15</sup> In these models, progression at 3- and 6-months were considered as binary variables. Prognostic variables included in the proportional hazards models were: treatment arm, age, gender, presence of bone metastases, presence of lung metastases.<sup>4,5</sup>

The association between PFS (denoted as  $x$ ) and OS (denoted as  $y$ ), was investigated on 1,381 mRCC patients using a non-parametric statistic that estimates Kendall's tau measure of association for bivariate time to event outcomes subject to censoring.<sup>16,17</sup> Kendall's tau measures the dependence between the two endpoints. For any pair of patients,  $(x_1, y_1)$  and  $(x_2, y_2)$  are said to be *concordant* if the ranks for both PFS and OS agree, that is, if both  $PFS_1 > PFS_2$  and  $OS_1 > OS_2$  or if both  $PFS_1 < PFS_2$  and  $OS_1 < OS_2$ . Otherwise the pairs are considered to be discordant. Kendall's tau ranges from  $-1$  to  $1$ , with  $1$  representing perfect agreement between the two rankings of the endpoints. If the disagreement between the two rankings is perfect, the coefficient will have value of  $-1$ . We would expect the coefficient to be approximately zero if the PFS and OS endpoints are independent. The null sampling distribution of the test statistic was approximated using 10,000 permutation replicates. Along with the bootstrap bias and standard error (SE), an interval estimate (CL) of the association parameter was constructed using adjusted bootstrap percentile confidence limits based on 10,000 permutation replicates.<sup>18</sup>

In addition, the adjusted association, a measurement to assess the surrogacy of a time-to-event endpoint was computed.<sup>19</sup> The null hypothesis to be tested was  $\theta = 1$  vs. the alternative hypothesis that  $\theta > 1$ . The adjusted association is also a measure of correlation between two censored variables, however we have applied the adjusted association for two reasons. First, there is a definite ordering between PFS and OS, PFS time cannot be greater than OS time. The semi-competing risk paradigm fits properly for this type of asymmetrical data structure. The semi-competing risk paradigm refers to a variation of the competing risk problem where a terminal event (deaths) censors a non-terminal event, but not vice versa. Second, there has been a great interest in adjusting for the treatment effect in the validation of the surrogate endpoint. The adjusted association first estimates the dependency parameter within each arm and then aggregates the information by a weighted average. The larger association suggests a higher concordance among all comparable pairs. By concordance, we mean whenever a patient has large PFS time, his/her overall survival duration is large too.

S-plus statistical software (version 8.0, Insightful Corp, Seattle, WA) and R were used for the data analyses and all statistical tests were two-sided.

## Results

### Baseline Characteristics

1,381 patients enrolled on the CALGB and AVOREN trials were included in this analysis. Baseline clinical characteristics are presented in Table 1. There were no differences in baseline clinical variables between the training and the testing datasets regarding gender, presence of liver or lung metastases, and distribution of patients among risk groups as assigned by established criteria.<sup>12</sup> There were differences between the two trials with regard to age, time from diagnosis of metastatic disease to study entry, prior nephrectomy, proportion of patients with bone metastases, and performance status. Patients who were randomized to CALGB 90206 were slightly older (median=62 years vs. 61 years), had more bone metastases as common site (29% vs. 19%), and better performance status (ECOG 0 is 62% vs. 41%) compared to patients randomized to the AVOREN trial. Neither trial had imbalances between experimental and control arms.

### Overall Survival

The median follow-up time among surviving patients was 62 (95% CI= 61–63) and 39 months (95% CI= 38–40) on the training and testing datasets, respectively. A total of 704 and 599 progression events and 627 and 444 deaths were observed in the training and testing datasets, respectively. The median survival time was longer in the testing dataset (AVOREN

trial, median=22 months) compared to the training dataset (CALGB 90206, median =18 months,  $p=0.06$ ), but this difference was not statistically significant when adjusted for the stratification factors (Figure 1, stratified log-rank  $p$ -value=0.483).

### Progression-Free Survival as a Predictor of OS

In the training dataset, the median OS time among patients who did not experience disease progression at 3-months was 25 months (95% CI=22–29) compared to 6 months (95% CI=5–7,  $p$ -value <0.0001) in those patients who experienced disease progression at 3-months. In the testing dataset, the median overall survival times were 26 months (95% CI= 24 – 30) and 6 months (95% CI = 5 – 9,  $p$ -value<0.0001) in patients who did not experience, and patients who did experience, disease progression at 3-months, respectively. The Kaplan-Meier product-limit survival curves by PFS rate at 3-months are displayed in Figure 2a–2b.

In the training dataset, disease progression at 3-months predicted OS (Table 2). The adjusted hazard ratio (HR) for death was 2.6 (95% CI=2.1–3.1,  $p$ -value<0.001) for patients who progressed at 3-months compared to patients who did not progress at 3-months. In the testing dataset, the adjusted HR for death was 3.4 (95% CI=2.7 – 4.3,  $p$ -value <0.001).

Similar results were obtained with the use of PFS at the 6-month landmark. In the training set, the median survival time among patients who did not and did experience progression at 6-months were 30 months (95% CI=25–35) and 8 months (95% CI=6 – 10,  $p$ -value<0.0001, Figure 3a) respectively. In the testing dataset, the median overall survival times were 32 months (95% CI= 27–39) and 10 months (95% CI =8–13,  $p$ -value<0.0001, Figure 3b) in mRCC patients who did not and did experience PFS at 6-months, respectively.

PFS at 6-months also predicted OS in multivariable analysis (Table 3). Compared to patients who did not progress at 6-months, the adjusted HR for death for patients who progressed at 6-months was 2.8 (95% CI=2.3 -3.4,  $p$ -value<0.001) in the training set and 2.8 (95% CI=2.2 – 3.5,  $p$ -value<0.001) in the testing data set.

### Dependence between PFS and OS

In the training dataset, the estimated Kendall's tau was 0.53 (95% CL=0.49–0.56, Table 4). Moreover, there is strong evidence suggesting that OS and PFS are statistically associated ( $p$ -value<0.00001). In patients who were treated with BV plus IFNA, Kendall's tau between the OS and PFS was estimated to be 0.60 (95% CL=0.55–0.63,  $p$ -value <0.00001). In contrast, Kendall's tau was 0.47 (95% CL=0.40–0.52) in patients who were treated with IFNA. This difference was statistically significant. The results based on a complete-case analysis (that is uncensored cases) yielded similar results.

Using the testing dataset, Kendall's tau between OS and PFS was 0.50 (95% CL=0.45–0.53,  $p$ -value <0.00001). Similarly, a higher coefficient of Kendall's tau was observed in patients who were treated with BV plus IFNA (Kendall's tau=0.58) compared to patients who were treated with IFNA plus placebo (Kendall's tau = 0.41, Table 4).

### Surrogacy between PFS and OS

Using the training set, the adjusted association was 3.145; a highly significant measure ( $p$ -value<0.0001) which was also confirmed in the testing set. The adjusted association was 2.80, which was also highly significant( $p$ -value <0.0001) and indicates that PFS is a surrogate marker of overall survival.

## Discussion

In this analysis of 1,381 patients with mRCC, progression-free survival at 3- and 6-months strongly predicted overall survival. In the training dataset, CALGB 90206, the median overall survival time was statistically significantly longer among mRCC patients who did not experience progression at 3-months (25 months) compared to patients who did experience progression (6 months,  $p$ -value $<0.0001$ ) with an adjusted hazard ratio for death for patients who experienced progression at 3-months of 2.6. Similar results were observed in patients who did and did not experience progression by 6-months. These findings were initially confirmed using the 649 mRCC patients randomized on the testing dataset, the AVOREN trial. It is noteworthy that similar results were also reported by Heng and colleagues in a retrospective analysis of 1,158 mRCC patients treated in 12 centers.<sup>11</sup>

In this current analysis, the relationship between PFS and OS was moderate with Kendall's tau of 0.53 and 0.50, respectively. However, the -relationship was highly statistically significant with  $p<0.00001$ . In fact a larger coefficient of Kendall's tau (0.60 and 0.58 in the CALGB and AVOREN trials) was observed in patients who were treated with bevacizumab plus IFNA compared to patients treated with interferon alpha alone (0.47 and 0.41 in the CALGB and AVOREN trials), indicating a stronger dependence between PFS and OS.

In addition, the adjusted association, a quantity to assess the surrogacy of PFS, demonstrates that PFS is a good surrogate of overall survival. Unlike Kendall's tau, the adjusted association is computed after incorporating the treatment assignment and it as expected for a good endpoint, the true association between the surrogate (PFS) and true endpoint (OS) improved after adjustment for the treatment effect. This association was confirmed in the testing set (AVOREN).

Prior to this analysis, the utility of PFS as a potential intermediate endpoint for OS in metastatic RCC has been called into question, insomuch as the PFS results of both CALGB 90206 and AVOREN, did not translate into a survival benefit. Further doubt about PFS as a useful intermediate endpoint was raised by the recently reported results of tivozanib versus sorafenib in which a PFS advantage in the experimental tivozanib arm was linked to an apparent survival improvement in the sorafenib arm.<sup>4-7</sup> These observations could be explained by a weak association between PFS and OS; although, an alternative explanation is that the association between these two endpoints is strong but was obscured by multiple sequential salvage therapies after progression. The latter explanation appears to be the case in the two analyzed bevacizumab trials as 62% and 54% of patients randomized on INFA or BV+INFA on CALGB 90206 received subsequent therapies. Although some investigators have made similar claims in the past,<sup>6,7</sup> this is the first formal statistical analysis of prospectively obtained clinical trial data supporting such a view and supports regulatory agency approval of bevacizumab as a useful therapy in metastatic RCC.

While the results of this analysis suggest that PFS may be used as a surrogate endpoint in trials with mRCC utilizing VEGF pathway directed therapies, it is important to mention that it was not possible to evaluate the PFS endpoint for surrogacy using the Prentice criteria as bevacizumab did not demonstrate a prolongation in overall survival in either trial.<sup>20</sup> Several authors, however, have used different approaches to the validation of surrogate endpoints (such as individual level surrogacy based on individual patient data).<sup>21-23</sup> Thus, an intermediate endpoint of OS may be of great clinical trial utility, even if it does not meet Prentice's criteria.

There are several strengths for this current analysis. First, the data were from randomized multi-institutional phase III trials that prospectively enrolled, treated patients and collected data of high quality. Second, both trials had similar eligibility criteria and recruited



previously-untreated mRCC patients to treatment with the same experimental regimen (bevacizumab and interferon alpha) and the same control arm was used (interferon alpha alone). Third, the results of this analysis were confirmed in an independent dataset. Finally, individual patient data were used in this analysis. The main limitation of this analysis is the fact that patients enrolled in these studies were required to have certain eligibility criteria deemed appropriate for participation in a clinical trial. Thus, the results of the study cannot be generalized to the entire mRCC patient population and may or may not be generalizable to other VEGF pathway or non-VEGF pathway directed therapies.

Nevertheless, there exists an unmet need for validated surrogate endpoints in mRCC clinical trials. While OS remains the regulatory gold standard, a pragmatic and reliable intermediate endpoint, especially in the era of multiple available therapies and prolonged overall survival, is critical. The current analysis demonstrate that PFS at 3- and at 6-months predicted OS in patients with mRCC treated with bevacizumab and interferon alpha and suggest that PFS may be an appropriate intermediate endpoint for OS. These findings though should be validated in mRCC patients treated with VEGF pathway directed and agents in general.

## Acknowledgments

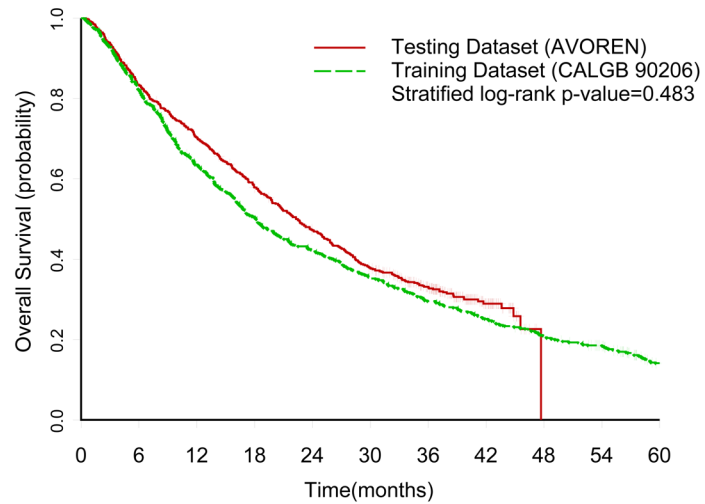
The authors would like to acknowledge Roche for sharing the AVOREN dataset. In addition, the authors thank Drs. Peter Compton, George Kong, and Nicola Moore for answering questions regarding the AVOREN database. Supported in part by the National Cancer Institute (grants CA031946 and CA033601 for CALGB 90206, and grant CA 155296-1A1).

## References

1. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med.* 2007; 356:115–124. [PubMed: 17215529]
2. Hudes G, Carducci M, Tomczak P, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med.* 2007; 356:2271–2281. [PubMed: 17538086]
3. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med.* 2007; 356:125–134. [PubMed: 17215530]
4. Rini BI, Halabi S, Rosenberg JE, et al. Bevacizumab plus interferon-alpha versus interferon-alpha monotherapy in patients with metastatic renal cell carcinoma: Results of CALGB 90206. *J Clin Oncol.* 2008; 26:5422–5428. [PubMed: 18936475]
5. Escudier B, Pluzanska A, Korawleski P, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet.* 2007; 370:2103–2111. [PubMed: 18156031]
6. Rini BI, Halabi S, Rosenberg JE, et al. A phase 3 study of bevacizumab plus interferon-alpha versus interferon-alpha monotherapy in patients with metastatic renal cell carcinoma-final results of CALGB 90206. *J Clin Oncol.* 2010; 28:2137–2143. [PubMed: 20368558]
7. Escudier B, Bellmunt J, Négrier S, et al. Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival. *J Clin Oncol.* 2010; 28:2144–2150. [PubMed: 20368553]
8. Halabi S, Vogelzang NJ, Ou SS, et al. Progression-free survival as a predictor of overall survival in men with castrate resistant prostate cancer (CRPC). *J Clin Oncol.* 2009; 27:2766–2771. [PubMed: 19380448]
9. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or nonsmall-cell lung cancer: A meta-analysis. *Lancet Oncol.* 2006; 7:703–704. [PubMed: 16945763]
10. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol.* 2007; 25:5218–5224. [PubMed: 18024867]

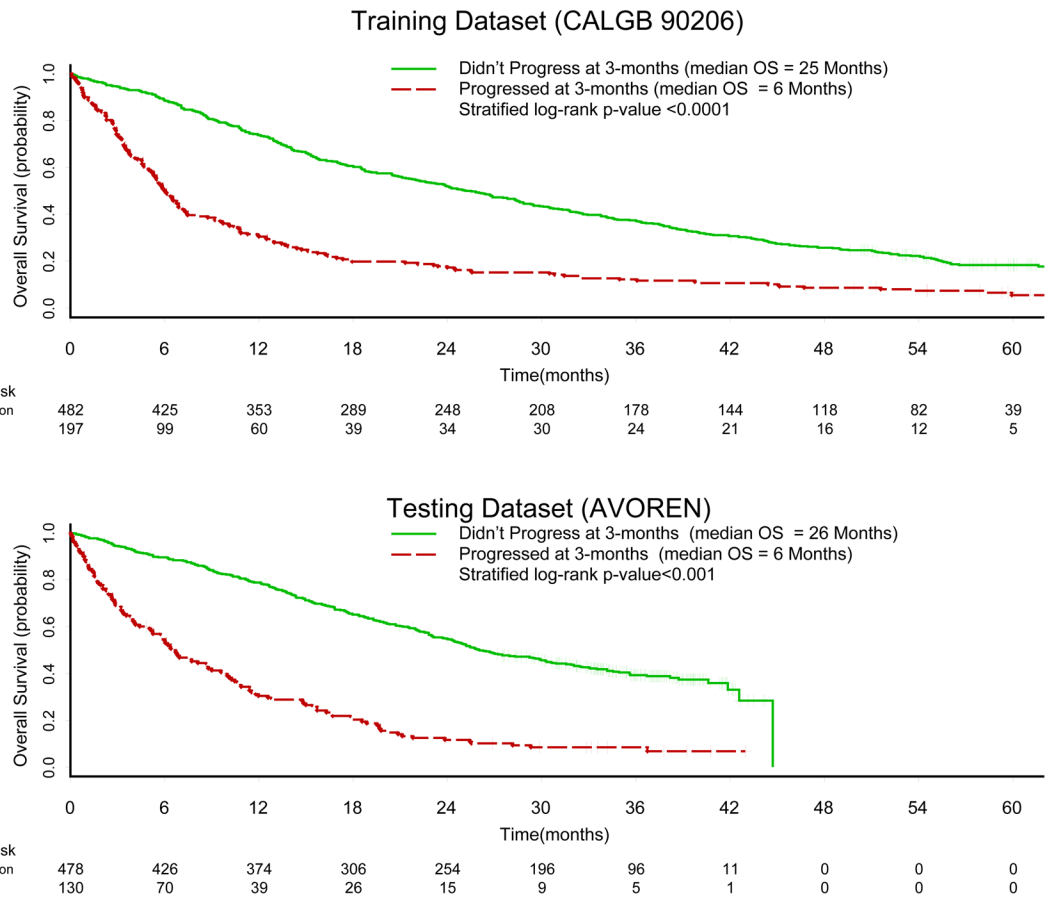
11. Heng DYC, Xie W, Bjarnason GA, et al. Progression-Free Survival as a Predictor of Overall Survival in Metastatic Renal Cell Carcinoma Treated With Contemporary Targeted Therapy. *Cancer*. 2011; 117:2637–2642. [PubMed: 21656741]
12. Motzer RJ, Mazumdar M, Bacik J, et al. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol*. 1999; 17:2530–2540. [PubMed: 10561319]
13. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958; 53:457–481.
14. Cox DR. Regression models and life tables. *J Roy Stat Soc*. 1972; B74:187–220.
15. Owzar K, Sen PK. Copulas: Concepts and novel applications. *Metron*. 2003; 61:1475–1489.
16. Akritas MG, Siebert J. A test for partial correlation with censored astronomical data. *Mon Not R Astron Soc*. 1996; 278:919–924.
17. Ghosh D. On assessing surrogacy in a single trial setting using a semicompeting risks paradigm. *Biometrics*. 2009; 65:521–529. [PubMed: 18759839]
18. Efron B. Better bootstrap confidence intervals (with discussion). *J Am Stat Assoc*. 1987; 82:171.
19. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*. 2008; 372:449–456. [PubMed: 18653228]
20. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989; 8:431–440. [PubMed: 2727467]
21. Freedman LS, Graubard BI, Schatzkin A, et al. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med*. 1992; 11:167–178. [PubMed: 1579756]
22. Buyse M, Molenberghs G, Burzykowski T, et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000; 1:49–67. [PubMed: 12933525]
23. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*. 1998; 54:1014–1029. [PubMed: 9840970]





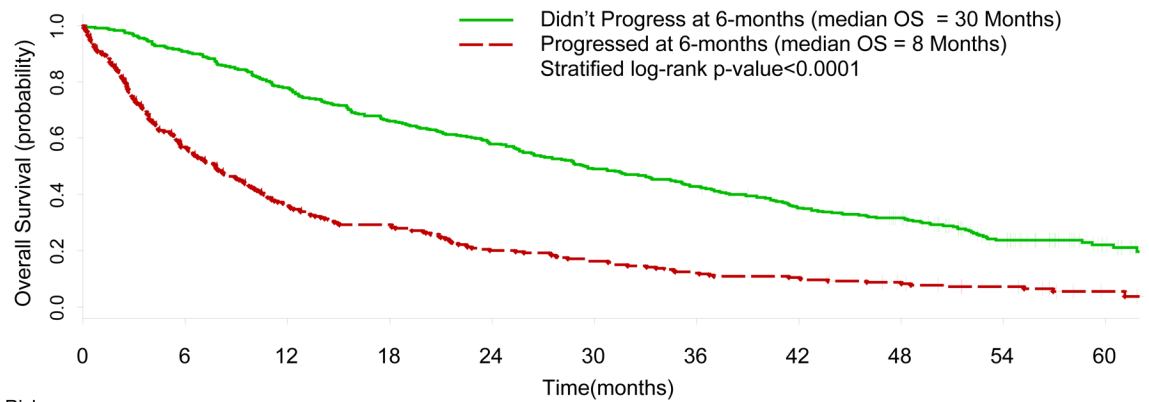
Number of Patients at Risk	
AVOREN	649 540 453 371 298 237 162 49 0 0 0
CALGB	732 600 462 366 307 256 215 181 148 122 66

**Figure 1.**  
 Overall Survival Kaplan-Meier Curves for the Training (CALGB 90206) and Testing (AVOREN) Datasets



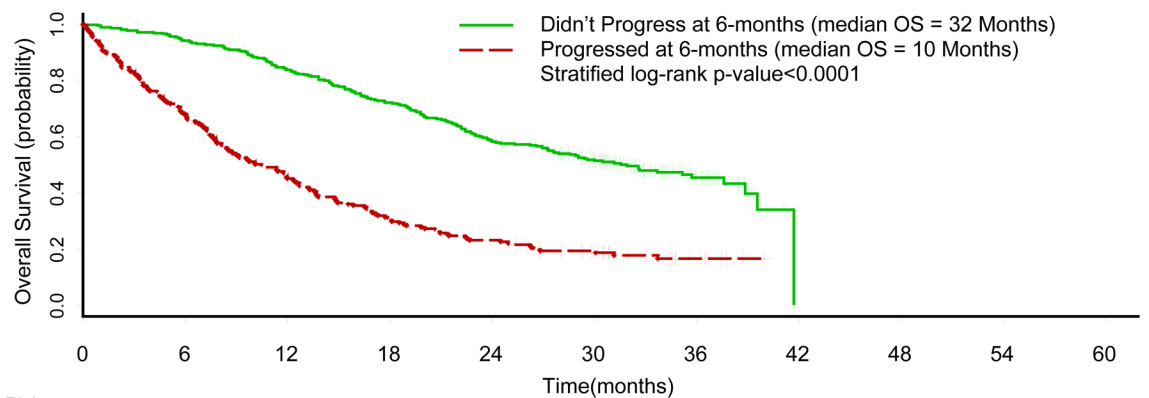
**Figure 2.** Overall Survival Kaplan-Meier Curves by PFS at 3 months in the Training (CALGB 90206) and Testing (AVOREN) Datasets.

## Training Dataset (CALGB 90206)



Num. Patients at Risk	0	6	12	18	24	30	36	42	48	54	60
Didn't Progress 6-mon	361	326	280	237	208	176	152	123	104	55	23
Progressed 6-mon	239	136	86	70	48	39	29	25	18	11	3

## Testing Dataset (AVOREN)



Num. Patients at Risk	0	6	12	18	24	30	36	42	48	54	60
Didn't Progress 6-mon	338	318	281	239	193	133	41	0	0	0	0
Progressed 6-mon	202	135	90	59	44	29	8	0	0	0	0

**Figure 3.** Overall Survival Kaplan-Meier Curves by PFS at 6-months in the Testing (CALGB 90206) and Testing (AVOREN) Datasets

**Table 1**

Baseline Clinical Characteristics of 1,381 Patients Randomized to CALGB 90206 and AVOREN Trials

Variable	Training Dataset (CALGB 90206, N=732)	Testing Dataset (AVOREN trial, N = 649)	Total N = 1,381
Gender			
Male	508 (69)	457 (70)	965 (70)
Female	224 (31)	192 (30)	416 (30)
Median Age, years (inter-quartile range)	62 (55.1 – 70.1)	60 (53 – 67)	61 (54 – 69)
Time since Metastatic Diagnosis, Months	3.1 (1.2 – 16.3)	2.9 (1.4 – 6.4)	3.0 (1.4 – 9.9)
Common Sites of Metastases *			
Lung	508 (69)	473 (73)	981 (71)
Bone	213 (29)	125 (19)	338 (24)
Liver	147 (20)	138 (21)	285 (21)
Measurable disease			
Yes	646 (90)	595 (92)	1241 (91)
No	75 (10)	54 (8)	129 (9)
Nephrectomy			
Yes	620 (85)	649 (100)	1269 (92)
No	112 (15)	0 (0)	112 (8)
ECOG Performance Status			
0	457 (62)	268 (41)	725 (52)
1	265 (36)	339 (54)	604 (44)
2	10 (1.4)	42 (5)	52 (4)
Number of adverse risk factors			
0 (favorable)	192 (26)	198 (31)	390 (28)
1–2 (intermediate)	465 (64)	392 (60)	857 (62)
3 (poor)	75 (10)	59 (9)	134 (10)
Laboratory			
Hemoglobin (g/dL)	13 (11.5 – 14.3)	13 (11.7 – 14.2)	13 (11.6 – 14.2)
LDH (U/KL)	1.7 (1.4 – 3.0)	2.6 (1.8 – 3.7)	2.8 (1.5 – 3.5)
Creatinine (mg/dl)	1.3 (1.1 – 1.5)	1.1 (0.9 – 1.2)	1.2 (1.0 – 1.4)
SGOT (U/L)	21 (16 – 26)	20 (15 – 25)	20 (16 – 26)
Bilirubin (mg/dL)	0.5 (0.3 – 0.7)	0.9 (0.7 – 1.2)	0.7 (0.4 – 1.0)
Albumin (g/dL)	4.0 (3.6 – 4.3)	4.1 (3.7 – 4.4)	4.0 (3.7 – 4.3)
Alkaline Phosphatase (U/L)	98(77–123.5)	101(77–164.5)	99 (77 – 134)
Arm			
BV + IFNA	369(50)	327 (50)	696 (50)
IFNA or Placebo	363 (50)	322 (50)	685 (50)

\* Not mutually exclusive

**Table 2**

Multivariable Proportional Hazards Models of PFS at 3-months Predicting Overall Survival for the Training and Testing Datasets

Factors	HR (95% CI, p-value) (CALGB 90206)	HR (95% CI, p-value) (AVOREN)
Disease Progression at 3-months (Yes vs. no)	2.6 (2.1–3.1, <0.0001)	3.4 (2.7–4.3, <0.0001)
Treatment Arm (BV+IFN vs IFN)	1.1 (0.9–1.3, 0.173)	1.1(0.9–1.3, 0.545)
Age (years)	1.0 (1.0–1.0, 0.653)	1.0(1.0–1.0, 0.800)
Gender (Females vs. males)	1.2 (1.0–1.5, 0.032)	1.1 (0.9–1.4, 0.322)
Presence of Bone Metastases (Yes vs. no)	1.1 (1.0–1.4, 0.147)	1.5 (1.2–1.9, 0.002)
Presence of Lung Metastases (Yes vs. no)	1.1 (0.9–1.4, 0.178)	1.1 (0.8–1.4, 0.577)

\* Adjusted for the stratification factors (prior nephrectomy, country, and number of adverse risk factors assessed using the MSKCC criteria)



**Table 3**

Multivariable Proportional Hazards Models of PFS at 6Months Predicting Overall Survival for the Training and Testing Datasets

Factors	HR (95% CI, p-value) (CALGB 90206)	HR (95% CI, p-value) (AVOREN)
Disease Progression at 6-months (Yes vs. no)	2.8 (2.3–3.4, <0.0001)	2.8 (2.2–3.5, <0.0001)
Treatment Arm (BV+IFN vs IF)	1.2 (1.0–1.5, 0.051)	1.1(0.9–1.4, 0.318)
Age (years)	1.0 (1.0–1.0, 0.476)	1.0(1.0–1.0, 0.828)
Gender (Females vs. Males)	1.3 (1.0–1.5, 0.020)	1.1 (0.9–1.5, 0.281)
Presence of Bone Metastases (Yes vs. no)	1.1 (0.9–1.3, 0.554)	1.5 (1.1–1.9, 0.007)
Presence of Lung Metastases (Yes vs. no)	1.1 (0.9–1.4, 0.336)	1.1 (0.8–1.4, 0.456)

\* Adjusted for the stratification factors (prior nephrectomy, country, and number of adverse risk factors assessed using the MSKCC criteria)

**Table 4**

Kendall's Tau (95% Confidence Limits) for the Dependence between PFS and OS Endpoints for the Training and Testing Datasets

Arm	Training Dataset (CALGB 90206) Kendall's tau (95% CL)	Testing Dataset (AVOREN) Kendall's tau (95% CL)
BV + IFNA	0.60 (0.55–0.63)	0.58 (0.52–0.62)
Placebo/IFNA	0.47 (0.40–0.52)	0.41 (0.33–0.48)
Total	0.53 (0.49–0.56)	0.50 (0.45–0.53)