CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

22-334

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES-TEAM LEADER'S MEMO

NDA/Serial Number:	22-334 / S-000
Drug Name:	Afinitor® (everolimus)
Indication(s):	Treatment of advanced renal cell carcinoma after prior treatment with sorafenib, sunitinib b(4)
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submitted: June 30, 2008 PDUFA: March 30, 2009 Review Completed: February 25, 2009
Review Priority:	Priority
Biometrics Division:	Division of Biometrics V (HFD-711)
Primary Statistical Reviewer:	Somesh Chattopadhyay, Ph.D.
Secondary Reviewers:	Shenghui Tang, Ph.D., Acting Team Leader
Concurring Reviewers:	Rajeshwari Sridhara, Ph.D., Deputy Director
Medical Division:	Division of Drug Oncology Products (HFD-150)
Clinical Team:	Qin Ryan, M.D., Medical Reviewer Robert Justice, M.D., Director
Project Manager:	Ms. Christy Cottrell

Keywords: Closed test procedures, Cochran-Mantel-Haenszel, data monitoring committee, double-blind, intent-to-treat, interim analysis, Kaplan-Meier product limit, log-rank test, multiple endpoints, proportional hazards, randomization, stratification, survival analysis.

The applicant has submitted results from one phase III, randomized, double-blind, comparative clinical trial (Study C2240) comparing Afinitor (everolimus or RAD001) plus best supportive care (BSC) and placebo plus BSC in patients with advanced renal cell carcinoma (RCC) who were previously treated with sunitinib, sorafenib or both sequentially. The trial started (first patient was dosed) on December 6, 2006. The trial stopped on February 28, 2008 because of improved PFS based on the results from an interim analysis with data cut-off date October 15, 2007. Updated data with cut-off date February 28, 2008 were subsequently submitted.

Study C2240 was a multicenter, international, randomized, double-blind, placebo-controlled, stratified, phase III trial. The study randomized 416 patients at 87 centers in 10 countries. Patients were randomized in a 2:1 proportion to receive either RAD001 at a dose of 10 mg (2 x 5 mg tablets) daily continuously or placebo. Patients were stratified at randomization according to: (1) Memorial Sloan Kettering Cancer Center (MSKCC) Risk Criteria (favorable vs. intermediate vs. poor) and (2) Number of VEGFr-TKI therapies taken by the patient prior to study entry (1 vs. 2).

The primary efficacy endpoint was progression-free survival (PFS). The secondary efficacy endpoints were overall survival (OS), overall response rate (ORR), and time to definitive deterioration in Functional Assessment of Cancer Therapy Kidney Symptom Index - Disease Related Symptoms (FKSI-DRS) and the physical functioning (PF) and quality of life (QL) scales of European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30.

For further details regarding the design, data analyses, and results of Study C2240, please refer to the statistical review by Dr. Somesh Chattopadhyay (March 18, 2009).

This study randomized a total of 416 patients, 277 to RAD001 arm and 139 to placebo arm. The RAD001 arm showed statistically significant improvement over placebo with respect to PFS as determined by independent radiologic review [hazard ratio=0.338, 95% confidence interval: (0.262, 0.436), log-rank test, p-value<0.0001 one sided and adjusted for MSKCC risk groups] in the intent-to-treat (ITT) population at the second interim analysis (significance level 0.0057) and also at the time when the study stopped (significance level 0.0193). The secondary endpoints OS and ORR were not significantly different between the two arms in the ITT population [hazard ratio=0.821, 95% confidence interval: (0.575, 1.171), log-rank test, p-value=0.138 one-sided and adjusted for MSKCC risk group for OS and Cochran-Mantel-Haenszel test stratified by MSKCC risk group, p-value=0.113, 2% in RAD001 vs. 0% in Placebo for ORR].

This Team Leader concurs with the recommendations and conclusions of the statistical reviewer (Dr. Somesh Chattopadhyay) of this application. The study showed benefit of RAD001 over placebo in terms of progression-free survival (PFS) as determined by independent radiologic review in this patient population based on the data from a planned interim analysis. However, the overall survival, a secondary endpoint, was not improved with RAD001 with approximately 32% overall deaths, but a trend favoring RAD001 was observed. RAD001 also did not show statistically significant superiority over placebo in

terms of overall response rate (another secondary endpoint) as determined by independent radiologic review. The data and statistical results provide adequate evidence to support the claims about PFS proposed in the NDA.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Shenghui Tang
3/18/2009 10:19:33 AM
BIOMETRICS

Rajeshwari Sridhara
3/18/2009 10:47:52 AM
BIOMETRICS

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	22-334 / S-000
Drug Name:	Afinitor® (everolimus)
Indication(s):	Treatment of advanced renal cell carcinoma after prior treatment with sorafenib, sunitinib b(4)
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submitted: June 30, 2008
	PDUFA: March 30, 2009
	Review Completed: March 18, 2009
Review Priority:	Priority
Biometrics Division:	Division of Biometrics V (HFD-711)
Statistical Reviewer:	Somesh Chattopadhyay, Ph.D.
Concurring Reviewers:	Shenghui Tang, Ph.D., Acting Team Leader
	Rajeshwari Sridhara, Ph.D., Deputy Director
Medical Division:	Division of Drug Oncology Products (HFD-150)
Clinical Team:	Qin Ryan, M.D., Medical Reviewer
	Robert Justice, M.D., Director
Project Manager:	Ms. Christy Cottrell

Keywords: Closed test procedures, Cochran-Mantel-Haenszel, data monitoring committee, double-blind, intent-to-treat, interim analysis, Kaplan-Meier product limit, logrank test, multiple endpoints, proportional hazards, randomization, stratification, survival analysis.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

1.1. Conclusions and Recommendations

The applicant has submitted results from one phase III, randomized, double-blind, comparative clinical trial (Study C2240) comparing Afinitor (everolimus or RAD001) plus best supportive care (BSC) and placebo plus BSC in patients with advanced renal cell carcinoma (RCC) who were previously treated with sunitinib, sorafenib or both sequentially. The study showed benefit of RAD001 over placebo in terms of progression-free survival (PFS) as determined by independent radiologic review in this patient population based on the data from a planned interim analysis. However, the overall survival, a secondary endpoint, was not improved with RAD001 with approximately 32% overall deaths, but a trend favoring RAD001 was observed. RAD001 also did not show statistically significant superiority over placebo in terms of overall response rate (another secondary endpoint) as determined by independent radiologic review. The data and statistical results provide adequate evidence to support the claims about PFS proposed in the NDA.

1.2. Brief Overview of Clinical Studies

The applicant has submitted results from one phase III, randomized, double-blind, comparative clinical trial (Study C2240) comparing Afinitor (everolimus or RAD001) plus best supportive care (BSC) and placebo plus BSC in patients with advanced renal cell carcinoma (RCC) who were previously treated with sunitinib, sorafenib or both sequentially. The trial started (first patient was dosed) on December 6, 2006. The trial stopped on February 28, 2008 because of improved PFS based on the results from an interim analysis with data cut-off date October 15, 2007. Updated data with cut-off date February 28, 2008 were subsequently submitted.

Study C2240 was a multicenter, international, randomized, double-blind, placebo-controlled, stratified, phase III trial. The study randomized 416 patients at 87 centers in 10 countries. Patients were randomized in a 2:1 proportion to receive either RAD001 at a dose of 10 mg (2 x 5 mg tablets) daily continuously or placebo. Patients were stratified at randomization according to: (1) Memorial Sloan Kettering Cancer Center (MSKCC) Risk Criteria (favorable vs. intermediate vs. poor) and (2) Number of VEGFr-TKI therapies taken by the patient prior to study entry (1 vs. 2).

The primary efficacy endpoint was progression-free survival (PFS). The secondary efficacy endpoints were overall survival (OS), overall response rate (ORR), and time to definitive deterioration in Functional Assessment of Cancer Therapy Kidney Symptom Index - Disease Related Symptoms (FKSI-DRS) and the physical functioning (PF) and quality of life (QL) scales of European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30.

1.3. Statistical Issues and Findings

Statistical Issues:

1. The quality of the datasets and the whole submission is poor. There were many mistakes and inconsistencies in the dataset that were revealed only after requests for clarification by the reviewer. Many documents were incomplete or not present at all. Some of them were submitted after the reviewers identified and requested them. Because of the poor quality of the datasets, inaccuracy of the results is a concern. For details of FDA's correspondence with the applicant, please refer to Appendix A.
2. Consideration of PFS as the primary endpoint for demonstration of efficacy for approval of drug products is based on the magnitude of the effect and the risk-benefit profile of the drug product. Because documentation of PFS assessments are often based on both subjective and objective criteria and these assessments depend on frequency, accuracy, reproducibility and completeness of tumor assessments, it is important that the observed magnitude of effect is robust. An interim PFS analysis may not provide an accurate or reproducible estimate of the treatment effect size due to inadequate follow-up, missing assessments, disagreements between radiological reviewers and/or disagreements between investigator and independent assessments. Stopping a trial based on interim PFS results which may not be verifiable after adjudication can be problematic and the trial results, in particular, may not be interpretable if the treatment in the control group was changed based on the interim results. In this application there was a substantial amount of disagreement between the independent and investigator's assessments of PFS. In addition, the design allowed the patients in placebo arm to crossover to RAD001 arm upon progression as assessed by the investigator. Although the effect of RAD001 on PFS is large, the robustness of this effect is in question. Nevertheless, since the updated data had approximately 91.7% of the events that were required for the final analysis, the lack of robustness of the PFS results is expected to be of a lesser degree.
3. There was disagreement in type or time of PFS event and censoring between independent and investigator's assessment for a large number of patients (56.0% in RAD001 arm and 39.6% in placebo arm. However the analysis results were similar based on these two assessments. The disagreement between independent and investigator's assessment seems more prominent in the RAD001 arm than in the placebo arm. This indicates a possible bias in investigator's assessment favoring RAD001 in spite of the study being double-blind.
4. Since the patients receiving placebo were crossed over to RAD001 on disease progression as assessed by the investigator and also when the study is terminated due to superior efficacy finding at the second interim analysis, it is difficult to evaluate the true effect of RAD001 on survival.

Findings:

This study randomized a total of 416 patients, 277 to RAD001 arm and 139 to placebo arm. The RAD001 arm showed statistically significant improvement over placebo with respect to PFS as determined by independent radiologic review [hazard ratio=0.338, 95% confidence interval:

(0.262, 0.436), log-rank test, p-value<0.0001 one sided and adjusted for MSKCC risk groups] in the intent-to-treat (ITT) population at the second interim analysis (significance level 0.0057) and also at the time when the study stopped (significance level 0.0193). The secondary endpoints OS and ORR were not significantly different between the two arms in the ITT population [hazard ratio=0.821, 95% confidence interval: (0.575, 1.171), log-rank test, p-value=0.138 one-sided and adjusted for MSKCC risk group for OS and Cochran-Mantel-Haenszel test stratified by MSKCC risk group, p-value=0.113, 2% in RAD001 vs. 0% in Placebo for ORR]. The analyses of the primary endpoint PFS and the secondary endpoints OS and ORR are presented in Table 1, Table 2 and Table 3, respectively.

Table 1: Primary Efficacy Analysis of PFS Based on Independent Review (Feb 28, 2008 Cut-off)

Treatment	Number of Patients	Number (%) Failed, Died, Progressed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	277	155 (55.96%) 21 (7.58%) 134 (48.38%)	4.90 (3.98, 5.52)	0.338 (0.262, 0.436)	<0.0001
Placebo	139	111 (79.86%) 8 (5.76%) 103 (74.10%)	1.87 (1.84, 1.94)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis.

Table 2: Analysis of OS (Feb 28, 2008 Cut-off)

Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	277	85 (30.69%)	NE (11.43, NE)	0.821 (0.575, 1.171)	0.138
Placebo	139	48 (34.53%)	13.01 (10.09, NE)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis. NE: Not estimable.

Table 3: Best Overall Tumor Response Based on Independent Review (Feb 28, 2008 Cut-off)

	Randomization Group		
	RAD001 (N=277)	**Placebo (N=139)**	**All (N=416)**
Complete Response	0	0	0
Partial Response	5 (1.81%)	0	5 (1.20%)
Stable Disease	185 (66.79%)	45 (32.37%)	230 (55.29%)
Progressive Disease	57 (20.58%)	74 (53.24%)	131 (31.49%)
Unknown	30 (10.83%)	20 (14.39%)	50 (12.02%)
Overall Response Rate (95% CI)	1.81% (0.59%, 4.16%)	0%	1.20%
P-value[1]	0.113		

[1]: Based on Cochran-Mantel Haenszel test stratified by MSKCC risk group.

2. INTRODUCTION

2.1. Overview

Renal cell carcinoma (RCC) is a form of kidney cancer that originates in the renal cortex and accounts for 85-90% of all kidney tumors.

RCC represents approximately 2% of all adult malignancies. More than 208,000 cases of kidney cancer are diagnosed per annum worldwide, including an anticipated 54,390 cases in the United States in 2008. In Europe, the highest incidence is observed in eastern European countries and in Germany. Clear cell is the most common histological subtype, accounting for approximately 75-80% of the cases. The incidence of RCC has increased annually by an estimated 2% over the past 2 decades and by 126% over the past 50 years. RCC is currently the sixth leading cause of cancer death and is responsible for >100,000 deaths worldwide each year (13,000 of which is estimated to be in the US). Chemotherapy and radiation therapy demonstrate little efficacy in advanced disease, and the 5-year survival rate has been low, ranging from 5-10%.

Approximately 25-30% of patients present with metastatic RCC (mRCC) at the time of the initial diagnosis. RCC affects more males than females (ratio of 1.6:1). The median age at diagnosis is 65 years, while that at death is 71 years.

2.1.1. Background

Everolimus (RAD001), a derivative of rapamycin, was initially developed to prevent allograft rejection following solid organ transplantation, and is approved in more than 60 countries worldwide for use in this indication (trade name Certican®).

Everolimus is being investigated as an anticancer agent based on its potential to act

- directly on the tumor cells by inhibiting tumor cell growth and proliferation,
- indirectly by inhibiting angiogenesis leading to reduced tumor vascularity (via potent inhibition of tumor cell VEGF production and VEGF-induced proliferation of endothelial cells, fibroblasts, and blood vessel-associated smooth muscle cells).

Everolimus is a derivative of rapamycin and acts as a signal transduction inhibitor. Its target is mTOR, a key regulatory serine-threonine kinase regulating metabolism, cell growth and proliferation, and angiogenesis. For treatment of RCC, the applicant reports that the most important action of everolimus is thought to be the inhibition of proangiogenic pathways at target sites distinct from the VEGFr protein tyrosine kinase (the target of sorafenib, sunitinib, and bevacizumab). The applicant further reports that inhibition of the mTOR kinase by everolimus leads to decreased protein translation including decreased production of hypoxia-inducible factor (HIF)-1a. In turn, decreased HIF-1a results in reduced secretion of angiogenic factors such as VEGF and fibroblast growth factor by the tumor. Moreover, proliferation of the endothelial cell, fibroblast, and smooth muscle cells required for blood vessel formation is inhibited by everolimus acting downstream of the angiogenic growth factor receptors. In addition to the dual

action of everolimus on tumor angiogenesis, it also acts directly to inhibit tumor cell growth and proliferation.

2.1.2. Regulatory History

Everolimus has been in clinical development as an investigational immunosuppressant drug for transplantation under IND [redacted] since 1996. Two NDAs for everolimus have been previously submitted on December 19, 2002 by Novartis Pharmaceutical Corp. for use in transplant patients, NDA [redacted] for the prophylaxis of organ rejection in allogenic kidney transplantation and NDA 21-628 for the prophylaxis of organ rejection in cardiac transplantation. The applications have not been approved because a safe regimen was not established for everolimus when used with cyclosporine. The FDA sent an approvable letter On October 20, 2003 to which the applicant submitted a complete response on February 27, 2004. The FDA sent another approvable letter on August 27, 2004 again citing the same reasons as were in the original action letter for not approving the application. The FDA has provisionally approved the trade name Certican for this drug under the IND phase of development. b(4)

Since November 2002, everolimus has also been in development to treat cancer patients both as monotherapy [redacted]; under IND 66,279. The dose regimen and tablet strengths used in the transplant setting are different from that in the oncology setting. The applicant submitted a special protocol assessment (SPA) request for the pivotal study C2240 on July 28, 2006. There was no agreement on the SPA. b(4)

2.1.3. Specific Studies Reviewed

This application is based on a single study C2240. This study was a multicenter, international, randomized, double-blind, placebo-controlled, stratified, phase III study designed to evaluate the safety and efficacy of RAD001 in patients with metastatic carcinoma of the kidney whose disease had progressed on Vascular Endothelial Growth Factor-receptor (VEGFr) tyrosine kinase inhibitor (TKI) therapy. The study started (first patient dosed) on 6 Dec 2006 and completed on 28 Feb 2008. Initial data cut-off date was 15 Oct 2007, which was used for the second interim analysis. By that time 410 patients were randomized, 272 to RAD001 and 138 to placebo. By the end of the study, a total of 416 subjects were randomized, 279 to RAD001 and 139 to placebo. Of the 416 randomized subjects, 322 were men and 94 were women, 367 were White, and the median age was 61 years (age range: 27 to 85 years).

This multinational study enrolled patients at 87 centers from 10 countries: Australia (6 centers), Canada (7 centers), France (8 centers), Germany (5 centers), Italy (8 centers), Japan (14 centers), Netherlands (4 centers), Poland (4 centers), Spain (5 centers), and USA (26 centers). Majority of the subjects were enrolled outside USA. A total of 111 (26.68%) subjects, 68 (25.55%) in everolimus arm and 43 (30.94%) in placebo arm were enrolled in the US.

2.2. Data Sources

Data used for this review are from the electronic submissions dated June 30, 2008 and September 30, 2008. The paths are \\Cdsesub1\EVSPROD\NDA022334\0000\m5\datasets\rad001c2240 and \\Cdsesub1\EVSPROD\NDA022334\0011\m5\datasets\rad001c2240.

3. STATISTICAL EVALUATION

3.1. Evaluation of Efficacy

The applicant has submitted efficacy results from a single study (Study C2240) titled "A randomized, double-blind, placebo-controlled, multicenter phase III study to compare the safety and efficacy of RAD001 plus Best Supportive Care (BSC) versus BSC plus Placebo in patients with metastatic carcinoma of the kidney which has progressed on VEGF receptor tyrosine kinase inhibitor therapy."

3.1.1. Study Design

This study was a multicenter, international, randomized, double-blind, placebo-controlled, stratified, phase III study designed to evaluate the safety and efficacy of RAD001 in patients with metastatic carcinoma of the kidney whose disease had progressed on Vascular Endothelial Growth Factor-receptor (VEGFr) tyrosine kinase inhibitor (TKI) therapy.

Major inclusion criteria included at least 18 years of age, metastatic carcinoma with histological or cytological confirmation of clear-cell RCC, progression on or within 6 months of stopping treatment with a VEGFr TKI (sunitinib and/or sorafenib), presence of at least one measurable lesion at baseline and Karnofsky Performance Score $\geq$ 70%.

Patients who met the study entry criteria were randomized in a 2:1 proportion to one of the two arms:

- RAD001 plus Best Supportive Care (BSC)
- BSC plus Matching Placebo

The randomization was stratified by:

1. Memorial Sloan Kettering Cancer Center (MSKCC) Risk Criteria (favorable vs. intermediate vs. poor);
2. Number of VEGFr TKI therapies taken by the patient prior to study entry (1 vs. 2).

A block randomization method was used within the strata.

The study used 3 risk prognostic factors which are applicable to treated patients (patients who had their primary tumor removed by surgery) from the MSKCC Risk Criteria as follows:

- Low Karnofsky Performance Status (< 80%),
- Low hemoglobin (= 13 g/dL for males and = 11.5 g/dL for females), and
- High corrected serum calcium (= 10 mg/dL).

Based on the presence or absence of risk prognostic factors at study entry, patients were classified into 3 distinctive risk groups:

1. the favorable group had none of the risk factors noted above,
2. the intermediate group had one risk factor, and
3. the poor risk group had two or more risk factors.

The VEGFr TKIs considered were sunitinib and sorafenib.

Subjects were instructed to self-administer 2 tablets of RAD001 (5 mg each) or placebo orally daily.

There were up to five different phases in the study: screening/baseline, blinded treatment, open-label RAD001, follow-up and the extension portion of the study.

Screening/baseline Phase: Screening evaluations were conducted to determine if the patient met the study inclusion/exclusion criteria and were performed within 5 weeks of the first dose of the study drug. Tumor measurements obtained within 2 weeks of the first dose of the study medication were used as a baseline reference to determine the date of disease progression on blinded treatment.

Blinded treatment phase: Patients who met all inclusion and exclusion criteria were randomized to receive active RAD001 or its Matching Placebo. The first day of blinded treatment began on Day 1, Cycle 1. Each treatment Cycle lasted 28 days. There was no fixed duration of treatment, thus, patients were permitted to continue on blinded treatment until the occurrence of tumor progression determined by the local radiologist or until unacceptable toxicity, death or discontinuation from the study for any other reason. The assessment of disease progression was determined according to the RECIST (Response Evaluation Criteria in Solid Tumors) guidelines.

Open-label treatment phase: At the first occurrence of radiologically documented, disease progression according to RECIST guidelines, the investigator could unblind the patient and subsequently could offer treatment with active RAD001 (open-label) if the patient was receiving placebo treatment.

Follow-up phase: Patients who discontinued blinded or open-label treatment for any reason had a follow-up visit which was scheduled 28 days after the last dose of the study drug. During this visit, the occurrence of AEs and SAEs after the last dose of the study drug was documented. Patients, who discontinued from the study, were asked to provide additional tumor assessments until the start of new anticancer therapy.

Extension treatment phase: An extension phase was added to the study to allow patients, who had been responding to treatment, to continue receiving treatment with open-label RAD001 until the occurrence of disease progression, or discontinuation from the study for any reason, or until RAD001 became commercially available in their country.

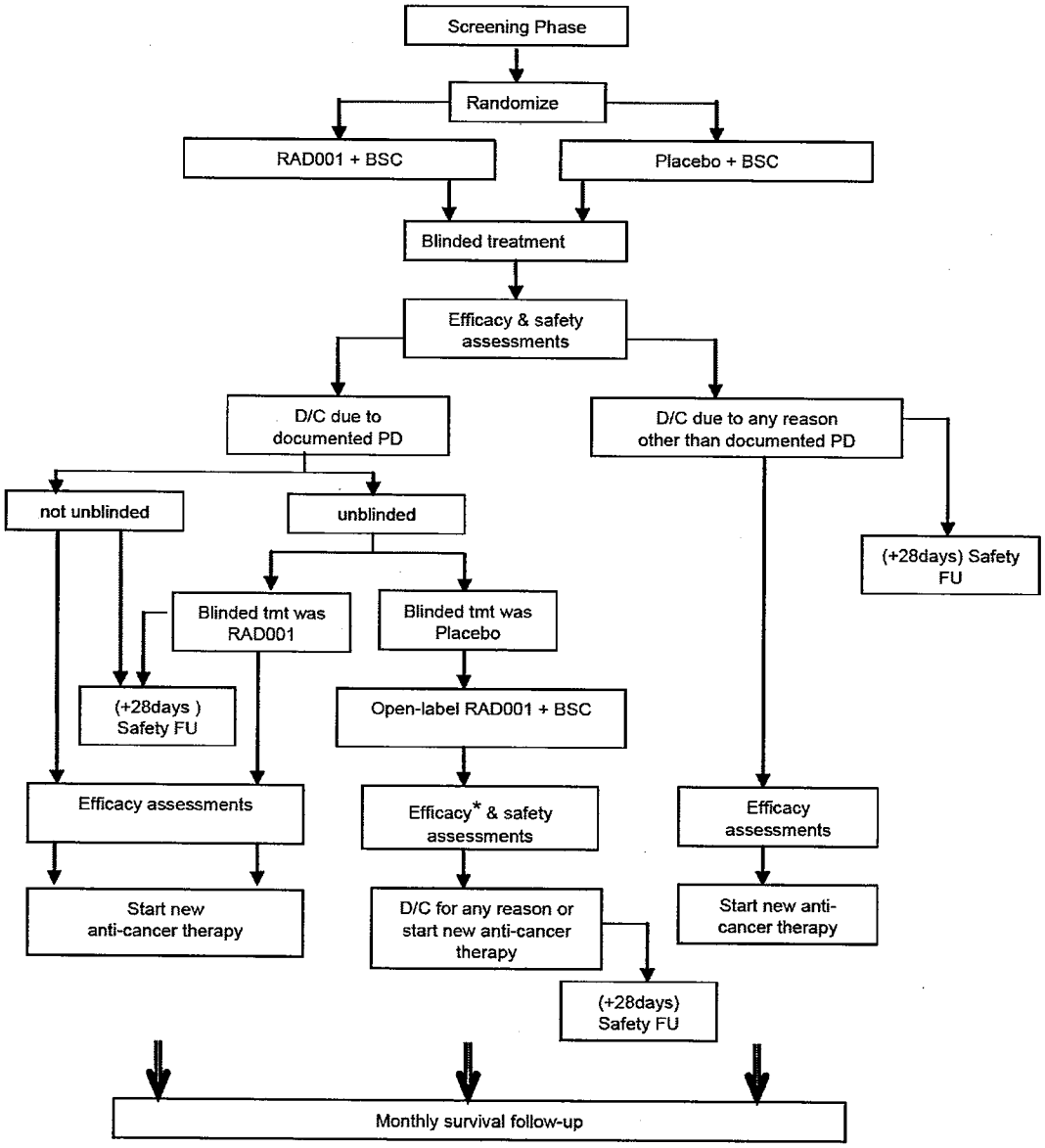
All patients receiving blinded and open-label treatment had routine safety and efficacy evaluations as follows:

Safety evaluations were performed on Day 1 of every treatment Cycle or as clinically indicated until discontinuation from the study. Efficacy evaluations were performed every 8 weeks (± 1 week) from the first dose of the study medication and included radiologic assessments (CT scans or MRIs, bone scans). All CT scans, MRIs, and bone scans obtained during the study and the

follow-up period were sent to an independent Central Radiologist whose assessment of disease progression was the basis for the primary analysis of the primary efficacy endpoint.

The study design flow chart is presented in Figure 1.

Figure 1: Study Design Flow Chart



* Scans collected in patients on open-label RAD001 should not be sent for central radiological review; PD= progressive disease; tmt = treatment; FU = follow-up.
Source: Clinical study report submitted with the NDA.

3.1.2. Schedule of Assessments

Tumor response and progression was assessed using the RECIST Criteria Guidelines. The assessment was done independently at the site (by a local radiologist) and by the central radiology review. A CT Scan of the Chest, Abdomen and Pelvis (CAP) obtained within 2 weeks of the first dose of the study medication was used as the baseline tumor assessment. Thereafter, tumor response was assessed every 8 weeks (±1 week) during the first year of treatment and every 12 weeks (±1 week) during and after the second year of treatment.

The methods of tumor assessment were CT Scans or MRIs with contrast. All lesions identified at baseline (target and non-target) were re-evaluated using the same method (CT scan or MRI) throughout the course of the study. Skin lesions selected as measurable disease (target lesions) had to be measured and color photographed.

After discontinuation from the study drug, all patients had monthly survival assessments. Survival assessments were to continue up to 2 years after the last patient was randomized.

The Functional Assessment of Cancer Therapy - Kidney Symptom Index, Disease Related Symptoms (FKSI-DRS) and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 questionnaires were used to assess patient reported outcomes. These patient questionnaires were administered on Day 1 of every treatment cycle and at discontinuation from the study drug to collect patient reported outcomes (PROs).

3.1.3. Efficacy Endpoints

Primary endpoint:

- Progression-free survival (PFS)

Secondary endpoints:

- Overall Survival (OS)
- Objective response rate (ORR)
- FKSI-DRS
- PF scale of EORTC QLQ-30
- QL scale of EORTC QLQ-30

Progression-free survival (PFS) is defined as the time from the date of randomization to the date of the first documented disease progression or death due to any cause. If a patient has not progressed or died at the date of the analysis cut-off or when he/she receives any further anti-neoplastic therapy (including open-label RAD001), PFS is censored at the time of the last tumor assessment before the cut-off date or the anti-neoplastic therapy date.
Overall survival (OS) is defined as the time from date of randomization to date of death due to any cause. If a patient is not known to have died, survival was to be censored at the date of last contact.

The overall response rate is defined as the proportion of patients with best overall response of complete response (CR) or partial response (PR) as determined by the independent central radiological review based on RECIST.

The duration of overall response (CR or PR) is defined as the time from first occurrence of PR or CR (as determined by the independent central radiological review) until the date of first documented disease progression or death due to any cause.

The Functional Assessment of Cancer Therapy - Kidney Symptom Index, Disease Related Symptoms (FKSI-DRS) and the European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30 questionnaires were used to assess patient reported outcomes.

The PRO variables of primary interest were the following.

1. Time to definitive deterioration of the FKSI-DRS by at least 2 units from baseline.
2. Time to definitive deterioration of the physical functioning (PF) scale of the EORTC QLQ-C30 by at least 10% frombaseline.
3. Time to definitive deterioration of the global health status / QoL (QL) scale of the EORTC QLQ-C30 by at least 10% frombaseline.

Definitive deterioration in FKSI-DSK is defined as a decrease in FKSI-DRS score by at least 2 units compared to baseline, with no later increase above this threshold observed during the course of the study. A single measure reporting a decrease of at least 2 units is considered definitive only if it is the last one available for the patient. Baseline is the latest available assessment made on or before the date of randomization. Time to definitive deterioration is the number of days between the date of randomization and the date of the assessment at which definitive deterioration is seen.

For patients dying before definitive deterioration, the definitive deterioration were to be considered to have occurred 5 weeks after the last available assessment if death occurs within 9 weeks from last assessment and definitive deterioration will be censored at the last assessment if death occurs more than 9 weeks after the last assessment.

If a definitive deterioration is observed after missing assessments this event was planned to be backdated to the last available assessment before the definitive deterioration plus 5 weeks. Patients that have not worsened as of the cutoff date were planned to be censored at the date of their last assessment before the cutoff.

Patients with a baseline score of 0 or 1 were to be excluded from this analysis, as will questionnaires with less than 5 out of the 9 questions answered.

Time to definitive deterioration in PF and QL scales of EORTC-QLQ-30 are defined and handled similarly.

3.1.4. Interim Analyses

Two interim analyses were planned in the protocol, one after approximately 30% (87 PFS events) and the other after approximately 60% (187 PFS events) of the targeted number of 290 PFS events required for the final statistical analysis. Both interim analyses were to allow for stopping for efficacy and for futility.

However, it was decided to not conduct formal testing of PFS at the first interim analysis because it was deemed premature to make meaningful changes to the further conduct of the study and unlikely to have an impact on the decision to randomize additional patients due to faster than expected enrollment. The first interim analyses focused only on safety comparisons including overall survival.

A Lan-DeMets α-spending function with O'Brien-Fleming type stopping boundaries is used for efficacy analyses. Futility analyses (β-spending function) are based on a rho-spending function with $\rho = 2.6$. For this specific β-spending function, the probability of stopping for futility is larger than 20% at the first and larger than 70% at the second interim analysis under the null hypothesis.

The nominal significance level for the final analysis was not affected by the futility stopping rule. The final efficacy analysis was to be based on the conservative approach to use the significance level according to the final O'Brien-Fleming type boundary obtained without futility stopping rule.

The cut-off date for the second interim analysis was set to 15 Oct 2007, determined using a statistical prediction model as when approximately 174 PFS events were to be observed. At this cut-off date occurrence of 191 out of targeted 290 PFS events (per central radiology review) was observed corresponding to an information fraction of 191/290 = 0.65862. The efficacy stopping boundary according to the protocol pre-specified O'Brien-Fleming type α-spending function for the second interim analysis was 2.5273 (Z-scale) corresponding to p=0.005747 on the p-value scale. The futility stopping boundary according to the pre-specified β-spending for the second interim analysis is 0.8374 (Z-scale) corresponding to p=0.2011 on the p-value scale.

After reviewing the results of the second interim analysis on 25 February 2008, the Independent Data Monitoring Committee recommended to stop the trial due to outstanding efficacy of RAD001 in terms of PFS.

Reviewer's Comments:

1. The trial stopped for the second interim analysis based on 191 PFS events (approximately 65.9% of total target events of 290) as determined by independent radiologic review. However, the updated data with cut-off date February 28, 2008 had 266 PFS (91.7%) events.
2. Consideration of PFS as the primary endpoint for demonstration of efficacy for approval of drug products is based on the magnitude of the effect and the risk-benefit profile of the

drug product. Because documentation of PFS assessments are often based on both subjective and objective criteria and these assessments depend on frequency, accuracy, reproducibility and completeness of tumor assessments, it is important that the observed magnitude of effect is robust. An interim PFS analysis may not provide an accurate or reproducible estimate of the treatment effect size due to inadequate follow-up, missing assessments, disagreements between radiological reviewers and/or disagreements between investigator and independent assessments. Stopping a trial based on interim PFS results which may not be verifiable after adjudication can be problematic and the trial results, in particular, may not be interpretable if the treatment in the control group was changed based on the interim results. In this application there was a substantial amount of disagreement between the independent and investigator's assessments of PFS. In addition, the design allowed the patients in placebo arm to crossover to RAD001 arm upon progression. Although the effect of RAD001 on PFS is large, the robustness of this effect is in question. Since the updated data had approximately 91.7% of the events that were required for the final analysis, the lack of robustness of the PFS results is expected to be of a lesser degree. However, according the design, the patients in placebo arm were allowed to crossover to RAD001 upon progression which makes it difficult to evaluate the survival effect even when the survival data become mature.

3. Initially no reports or results of the first interim analysis have been submitted in this application. The applicant provided the meeting minutes of the IDMC meeting for the first interim analysis upon FDA's request. The first interim analysis was performed on October 17, 2007 only for safety. The amended plan to not perform any PFS analysis at the first interim analysis was submitted on February 1, 2008 by which time the first interim analysis had already been conducted.
4. Initially no independent data monitoring committee charter and the report and meeting minutes of the data monitoring committee for the first interim analysis had been submitted in this application. Those are submitted late upon FDA's request.
5. Although the study stopped at the second interim analysis based on data cut-off date October 15, 2007, the updated PFS data with cut-off date February 28, 2008 could be formally analyzed since PFS was statistically significant at the second interim analysis. Even if PFS were not statistically significant at the second interim analysis, the updated data could be analyzed because of the use of a spending function approach as long as it did not cross the futility boundary at an interim analysis. Based on the α-spending functions and the observed number of events, the significance levels for data cut-off of October 15, 2007 and February 28, 2008 were 0.0057 and 0.0193, respectively.

3.1.5. Sample Size Considerations

For sample size calculation a one sided type I error of α=0.025 and power 1- β = 0.9 were used for a 3-look group sequential plan with a Lan-DeMets α-spending function of O'Brien-Fleming type stopping boundaries for efficacy and a β-spending function based on a rho-spending function with ρ = 2.6 for futility. Assuming a hazard ratio of 1.5 (corresponding to a median PFS of 3 months for the Placebo arm and 4.5 months for RAD001 arm), and using a 2:1 randomization to RAD001 vs. Placebo a total 290 events for PFS were required.

Considering a recruitment time of 16 months and an additional follow up of 5 months a total of 362 patients had to be included. This number included the assumption that about 10% of patients are lost to follow up during the study. The study actually enrolled 416 patients.

The final analysis was to be performed when approximately 290 PFS events as per independent central radiological review were observed in the intent-to treat (ITT) population.

3.1.6. Efficacy Analysis Methods

3.1.6.1. Analysis Populations

The primary population for all efficacy analyses was the full analysis set (FAS) which is commonly known as the intent-to-treat (ITT) population. This population was defined as all randomized subjects and was analyzed according to the treatment and stratum patients were assigned to at randomization. The safety population consisted of all patients who received at least one dose of study drug and who had at least one valid post-baseline safety assessment. Patients were analyzed according to the treatment actually received. The study did not use any per-protocol population.

3.1.6.2. Demographic and Baseline Characteristics

Demographic and baseline characteristics were summarized by means of contingency tables for each treatment group for qualitative data (gender, race, disease stage, Karnofsky performance status, etc.) and by appropriate descriptive statistics (mean, standard deviation, median, minimum, and maximum) for each treatment group for quantitative data (age, body weight, etc.). Diagnosis and extent of cancer, medical history and ongoing conditions, duration of study drug exposure, cumulative dose, dose intensity, relative dose intensity, concomitant medications and significant non-drug therapies were summarized.

3.1.6.3. Analysis of Primary Endpoint

The primary efficacy endpoint of this study was PFS. The primary analysis of PFS is based on the central radiological assessments. As a default censoring and event date options the PFS was censored at the last adequate tumor assessment if one of the following occurred: absence of event; the event occurred after a new anti-neoplastic therapy (including open-label RAD001) is given; the event occurred after two or more missing tumor assessments.

The primary statistical analysis to compare PFS was performed using a one-sided log-rank test stratified by the MSKCC risk category at a significance level of α=0.025. The plot of Kaplan-Meier estimate of PFS in each treatment group is displayed. The plots display the number of patients at risk at equidistant time points. Median PFS for each treatment group was obtained

along with 95% confidence intervals. Kaplan-Meier estimates with 95% confidence intervals at 4, 6 and 12 months were summarized. The hazard ratio of the treatment effect estimated using a Cox proportional hazard model stratified by the MSKCC risk category was provided with two-sided 95% confidence interval.

In addition to the final analysis of PFS at the end of the study, 2 formal interim analyses of PFS were planned. A Lan-DeMets alpha spending function for O'Brien-Fleming type boundaries was used to ensure that the false type I error rate for PFS is less than or equal to 0.025 (1-sided).

3.1.6.4. Analysis of Secondary Endpoints

As secondary efficacy variables overall survival, objective response rate and the duration of response were to be compared between treatment arms. Additionally, the following patient reported outcomes were used as secondary endpoints: Disease-Related Symptoms of the FKSI-DRS (FKSI), Physical functioning scale (PF) of the EORTC QLQ-C30 and Global health status / QoL scale (QL) scores of the EORTC QLQ-C30.

According to the statistical analysis plan, the secondary endpoints were to be tested using a gate keeping testing procedure (Hommel, et al, 2007) to adjust for multiple testing.

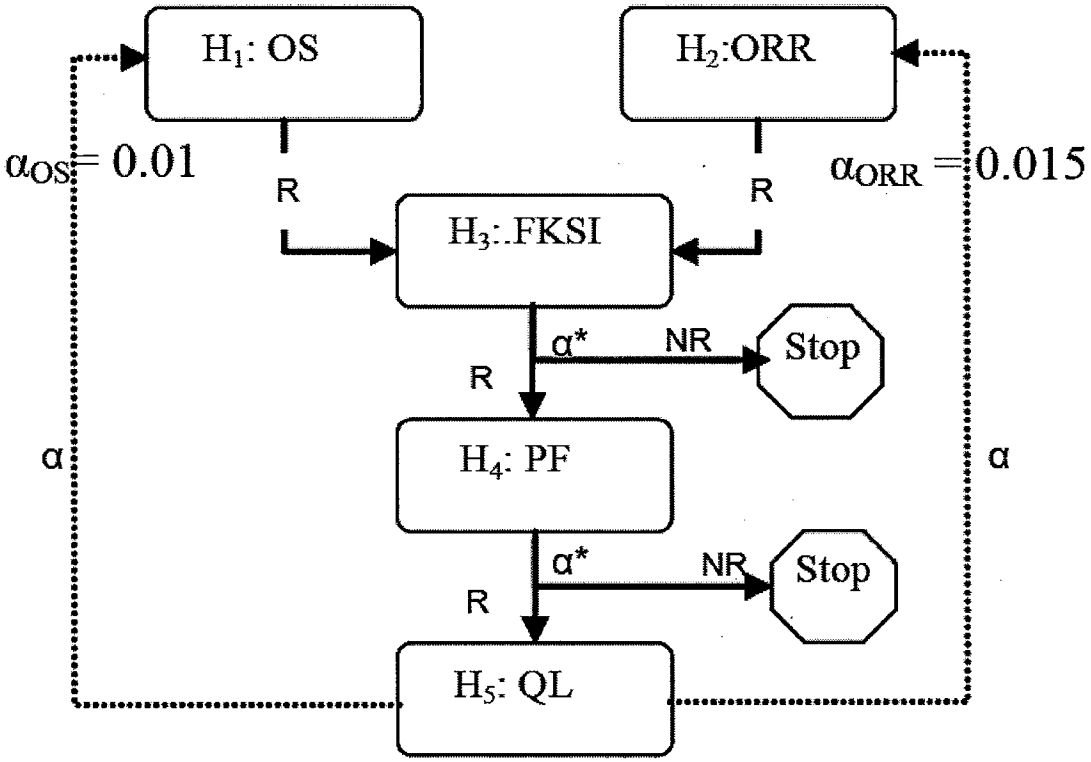
Hypotheses OS and ORR were to be tested separately at the one-sided 1% level and one-sided 1.5% level, respectively, using the Bonferroni approach.

The combination of the strategies below fulfils the general principle of the closed testing procedure.

1. If both OS and ORR hypotheses are rejected, then the a-priori ordered hypotheses for FKSI-DRS, PF scale of the EORTC QLQ-C30 and QL scale of the EORTC QLQ-C30 were be tested sequentially at the one-sided 2.5% level each and. In order to control the multiple type I error at 2.5% level, a hypothesis among the hypotheses for these three endpoints is rejected in a confirmatory way if and only if that hypothesis and all preceding hypotheses are rejected each at 2.5% level.
2. If only one of OS and ORR hypotheses is rejected then the a-priori ordered hypotheses FKSI-DRS, PF scale of the EORTC QLQ-C30 and QL scale of the EORTC QLQ-C30 were to be tested sequentially at the one-sided $\alpha^*=1\%$ or 1.5 % level ($\alpha^*=1\%$ if OS or $\alpha^*=1.5\%$ if ORR is rejected). In order to control the multiple type I error at α^* level, a hypothesis among the hypotheses for these three endpoints is rejected in a confirmatory way if and only if that hypothesis and all preceding hypotheses are rejected each at α^* level. If all three hypotheses are rejected each at α^* level, then the hypothesis that was not previously rejected (OS or ORR) was to be re-tested at the one-sided 2.5% level.
3. If none of OS and ORR hypotheses are rejected, the procedure stops.

This stepwise multiple test procedure controls the familywise error rate at the 2.5% level Strategies 1 and 2 are illustrated in Figure 2.

Figure 2: Gatekeeping Strategy for Multiple Testing of Secondary Endpoints



R= Hypothesis of no difference between treatment arms rejected,
NR= Hypothesis of no difference between treatment arms not rejected,
$\alpha^*=\alpha = 0.025$ (strategy 1) or $\alpha^* = 0.01$ or 0.015 (strategy 2)
Source: Statistical analysis plan submitted with the NDA

The gate keeping testing procedure strongly controls the overall significance level for multiple testing at 2.5%.

Since duration of response obtained for patients with a tumor response (CR or PR) only, the arms are not compared using a formal testing procedure. Therefore, duration of response is not implemented in the hierarchical testing procedure.

Since neither overall survival nor objective response rate met the criteria for significance (predefined in the gate keeping procedure), no formal testing of the PRO endpoints could be made.

Secondary endpoints included the following time-to event variables:

- Duration of response
- Overall survival

- Time to definitive deterioration of the FKSI-DRS score by at least 2 score units from baseline
- Time to definitive deterioration of the physical functioning scale (PF) score of the EORTC QLQ-C30 by at least 10% from baseline
- Time to definitive deterioration of the Global health status / QoL scale (QL) score of the EORTC QLQ-C30 by at least 10% from baseline

All time-to-event endpoints except duration of response (which was to be analyzed descriptively only) were analyzed similar to PFS analysis.

The overall response rate (ORR) was summarized in terms of percentage rates with 95% confidence intervals for each treatment group. ORR was compared between the treatment arms using stratified exact Cochran-Mantel-Haenszel test, using the strata defined by the MSKCC risk criteria. Duration of response was to be censored using the same rules as PFS.

3.1.6.5. Sensitivity Analyses

The following sensitivity and supportive analysis of PFS were performed:

- Using investigator assessments and the same conventions as for the primary analysis
- Using central radiology assessments and taking the PFS event whenever it occurs even after two or more missing tumor assessments.
- Using central radiology assessments with backdating of events occurring after missing tumor assessments.
- Combining both central radiology and investigator assessments and using the same conventions as for the primary analysis. If there was a PFS event or censoring for both central radiology and investigator then the time to event or censoring time is defined as earlier of those two times. If a PFS event was observed in only one of the sources, then the time of that PFS event was taken.
- Using a multivariate Cox model stratified by the MSKCC criteria, and adjusting for age, sex, prior treatment with sunitinib and sorafenib.

Reviewer's Comments:

1. Although the MSKCC risk category and the number of prior VEGFr-TKI therapies were used as stratification factor at randomization, number of prior VEGFr-TKI therapies was not used as a stratification factor for the analyses of primary and key secondary endpoints. Only the MSKC risk category was used as the stratification factor in those analyses. This plan was pre-specified in the protocol.
2. According to the second protocol amendment dated 28 Feb 2007, the secondary endpoints were to be tested according to a hierarchical testing procedure in the following order.
   - OS

- ORR
- Duration of response
- FKSI-DRS
- PF scale of EORTC QLQ-C30
- QL scale of EORTC QLQ-C30

The strategy was later modified in the Statistical Analysis Plan dated 1 Feb 2008 to introduce Hommel's gatekeeping procedure to adjust for testing of multiple secondary endpoints.

3. The validity of the patient-reported outcome instruments employed is questionable. Also, since neither OS nor ORR was statistically significant, the applicant did not analyze the patient-reported outcome following the gate keeping strategy. Therefore, this review considers those endpoints exploratory.
4. The study report included time to definitive worsening of Karnofsky performance status by at least one Karnofsky category (i.e. at least 10 points less) compared to baseline as a secondary endpoint. However, it was not specified in the protocol. Although the statistical analysis plan included the definition and analysis method for this endpoint, it was not categorized as a patient reported outcome variable.

3.1.7. Sponsor's Results and Statistical Reviewer's Findings/Comments

A total of 416 patients were randomized in study C2240. The study started (first patient dosed) on 6 Dec 2006.

The second interim analysis was conducted based on the data cut-off date October 15, 2007. After reviewing the results of the second interim analysis on February 25, 2008 the Independent Data Monitoring Committee (IDMC) recommended to stop the trial due to outstanding efficacy of RAD001 in terms of PFS. Novartis notified the investigators of this early stopping. All sites with patients receiving placebo were notified on February 28, 2008 to cross these patients over to RAD001.

At the initial submission the applicant submitted results based on data cut-off date that was used for the second interim analysis. The applicant later submitted the results based on data cut-off date of February 28, 2008. This review uses both datasets.

All analyses reported in this review are based on intent-to-treat (ITT) population which is also called "full analysis set" by the applicant.

3.1.7.1. Patient Disposition

At the time of the data cut-off for the second interim analysis (October 15, 2007) study 410 patients were randomized. At the time of the final data cut-off date (February 28, 2008) there were 416 patients randomized, 277 to RAD001 and 139 to placebo. Five patients, 3 in the RAD001 arm and 2 in the placebo arm, were randomized but did not receive treatment. No

patients received wrong treatment (treatment of the other arm than the one to which the patient is randomized). Patient disposition at the end of the study (February 28, 2008) based on ITT population is given in Table 4.

Table 4: Patient Disposition at the End of the Study (February 28, 2008) Based on ITT Population

Disposition	RAD001 N=277 (%)	Placebo N=139 (%)	Total N=416 (%)
Ongoing	75 (27.1)	6	81 (19.5)
Discontinued	202 (72,9)	133	335 (80.5)
Main reason for discontinuation			
Disease Progression	137 (49.5)	124 (89.2)	261 (62.7)
Death	7 (2.5)	4 (2.9)	11 (2.6)
Adverse event	36 (13.0)	2 (1.4)	38 (9.1)
Patient withdrew consent	13 (4.7)	2 (1.4)	15 (3.6)
Lost to follow-up	4 (1.4)	0 (0)	4 (1.0)
Protocol violation	2 (0.7)	1 (0.7)	3 (0.7)
Administrative problems	2 (0.7)	0 (0)	2 (0.5)
Abnormal laboratory values	1 (0.4)	0 (0)	1 (0.2)

3.1.7.2. Protocol Deviation

Overall, the number of eligibility criteria deviations was low. All patients had a diagnosis of mRCC, but five patients did not have a component of clear-cell carcinoma, three with papillary carcinoma and 2 with granular histology. The most common eligibility criteria deviation was patients having received concomitant immunosuppressive agents, such as chronic corticosteroids, at the time of study entry (about 4% in each arm).

3.1.7.3. Baseline Characteristics

The treatment arms were mostly comparable for demographic characteristics (gender, race and age) at baseline, although there was some imbalance between treatment arms in terms of age group, RAD001 having a higher percentage (40.4%) of elderly (≥65 years) patients than the placebo arm (29.5%). Most patients in the ITT population were male (77.4%). Also most of the patients in the ITT population were White (88.2%). The mean and median age were 60.2 years and 61 years, respectively; minimum and maximum were 27 and 85 years. Approximately 36.8% of patients were elderly. A summary of demographic characteristics at baseline is presented in Table 5.

Table 5: Demographic Characteristics: Gender, Race and Age (Feb 28, 2008 Cut-off)

		RAD001 (N=277)	Placebo (N=139)	All (N=416)
Gender	Female	61 (22.0%)	33 (23.7%)	94 (22.6%)
	Male	216 (78.0%)	106 (76.3%)	322 (77.4%)
Race	Asian	16 (5.8%)	11 (7.9%)	27 (6.5%)
	Black	2 (0.7%)	3 (2.2%)	5 (1.2%)
	Caucasian	246 (88.8%)	121 (87.1%)	367 (88.2%)
	Native American	1 (0.4%)	0 (0%)	1 (0.2%)
	Other	8 (2.9%)	3 (2.2%)	11 (2.6%)
	Missing	4 (1.4%)	1 (0.7%)	5 (1.2%)
Age Group in Years	<65	165 (59.6%)	98 (70.5%)	263 (63.2%)
	≥65	112 (40.4%)	41 (29.5%)	153 (36.8%)
Age in Years	Mean	60.66	59.27	60.20
	SD	10.36	9.58	10.11
	Minimum	27	29	27
	First quartile	54	54	54
	Median	61	60	61
	Third quartile	68	66	68
	Maximum	85	79	85

This study recruited patients from 10 countries. The countries are grouped by geographic regions as North America, Europe, and Australia and Japan. Majority of the randomized patients were from Europe region (60.34%). Highest number of patients were from USA (26.68%) followed by France (25.24%) and Italy (16.59%). All other countries each contributed less than 10% of patients. Two treatment arms were balanced for the most countries except USA (25.55% in RAD001 arm and 30.94% in placebo arm), Spain (6.50% vs. 2.88%) and Canada (3.25% vs. 7.19%). The distribution of country and region of the patients is presented in Table 6. Other baseline characteristics are presented in Table 7.

Table 6: Demographic Characteristics: Country and Geographic Region (Feb 28, 2008 Cut-off)

Geographic Region/Country	RAD001 (N=277)	Placebo (N=139)	All (N=416)
North America	77 (27.80%)	53 (38.13%)	130 (31.25%)
Canada	9 (3.25%)	10 (7.19%)	19 (4.57%)
USA	68 (25.55%)	43 (30.94%)	111 (26.68%)
Europe	180 (64.98%)	71 (51.08%)	251 (60.34%)
France	72 (25.99%)	33 (23.74%)	105 (25.24%)
Germany	14 (5.05%)	4 (2.88%)	18 (4.33%)
Italy	46 (16.61%)	23 (16.55%)	69 (16.59%)
Netherlands	11 (3.97%)	1 (0.72%)	12 (2.88%)
Poland	19 (6.86%)	6 (4.32%)	25 (6.01%)
Spain	18 (6.50%)	4 (2.88%)	22 (5.29%)
Australia and Japan	20 (7.22%)	15 (10.79%)	35 (8.41%)
Australia	5 (1.81%)	6 (4.32%)	11 (2.64%)
Japan	15 (5.42%)	9 (6.47%)	24 (5.77%)

Table 7: Baseline Non-demographic Characteristics (Feb 28, 2008 Cut-off)

		RAD001 (N=272)	Placebo (N=138)	All (N=410)
BMI (kg/m^2)	N	268	136	404
	Mean	26.31	26.22	26.28
	SD	5.01	4.33	4.79
	Median	25.6	25.3	25.5
	Minimum	15.9	20.8	15.9
	Maximum	47.5	40.2	47.5
Karnofsky Performance Status	100	78 (28.16%)	41 (29.50%)	119 (28.61%)
	90	98 (35.38%)	53 (38.13%)	151 (36.30%)
	80	72 (25.99%)	30 (21.58%)	102 (24.52%)
	70	28 (10.11%)	15 (10.79%)	43 (10.34%)
	Missing	1 (0.36%)	0 (0%)	1 (0.24%)
MSKCC Risk Group	Favorable risk	81 (29.24%)	39 (28.06%)	120 (28.85%)
	Intermediate risk	156 (56.32%)	79 (56.83%)	235 (56.49%)
	Poor risk	40 (14.44%)	21 (15.11%)	61 (14.66%)
Prior VEGFr-TKI Therapies	Sorafenib	81 (29.24%)	43 (30.94%)	124 (29.81%)
	Sunitinib	124 (44.77%)	60 (43.17%)	184 (44.23%)
	Both	72 (25.99%)	36 (25.90%)	108 (25.96%)

3.1.7.4. Primary Efficacy Analysis

Primary efficacy analysis comparing progression-free survival (PFS) based on independent radiologic review between RAD001 and placebo, in the ITT population using log-rank test adjusted for stratification factor baseline MSKCC risk group (favorable vs. intermediate vs. poor) is presented in Table 8 and Table 9 for data with October 15, 2007 and February 28 cut-off dates, respectively. The corresponding Kaplan-Meier plots are given in Figure 3 and Figure 4, respectively. The PFS analyses based on investigator's assessment are presented in Table 10 and Table 11 and Kaplan-Meier plots are presented in Figure 5 and Figure 6, respectively for these two cut-off dates. RAD001 demonstrated a statistically significant advantage in PFS over placebo (log-rank test, nominal one-sided p-value < 0.0001 adjusted for MSKCC risk group and not adjusted for interim analyses for both cut-off dates and for both PFS determined by independent assessment and PFS determined by investigator assessment). The hazard ratios of RAD001 over placebo and the 95% confidence intervals for PFS as determined by independent review were 0.306 [95% CI: (0.226, 0.413)] and 0.338 [95% CI: (0.262, 0.436)] based on data cut-off dates October 15, 2007 and February 28, 2008, respectively. The hazard ratios and the confidence intervals for PFS based on investigator's assessment were 0.314 [95% CI: (0.238, 0.414)] and 0.326 [95% CI: (0.256, 0.414)] for those two cut-off dates. The efficacy stopping boundary according to the protocol pre-specified O'Brien-Fleming type α-spending function for the interim analysis based on October 15, 2007 data cut-off with 191 events (approximately 65.86% of total events) was 0.005747 on the p-value scale which was crossed.

Table 8: Primary Efficacy Analysis of PFS Based on Independent Review (Oct 15, 2007 Cut-off)

Treatment	Number of Patients	Number (%) Failed, Died, Progressed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	272	101 (37.13%) 16 (5.88%) 85 (31.25%)	4.01 (3.71, 5.52)	0.306 (0.226, 0.413)	<0.0001
Placebo	138	90 (65.22%) 8 (5.80%) 82 (59.42%)	1.87 (1.81, 1.94)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis.

Figure 3: Kaplan-Meier Plot of PFS Based on Independent Review (Oct 15, 2007 Cut-off)

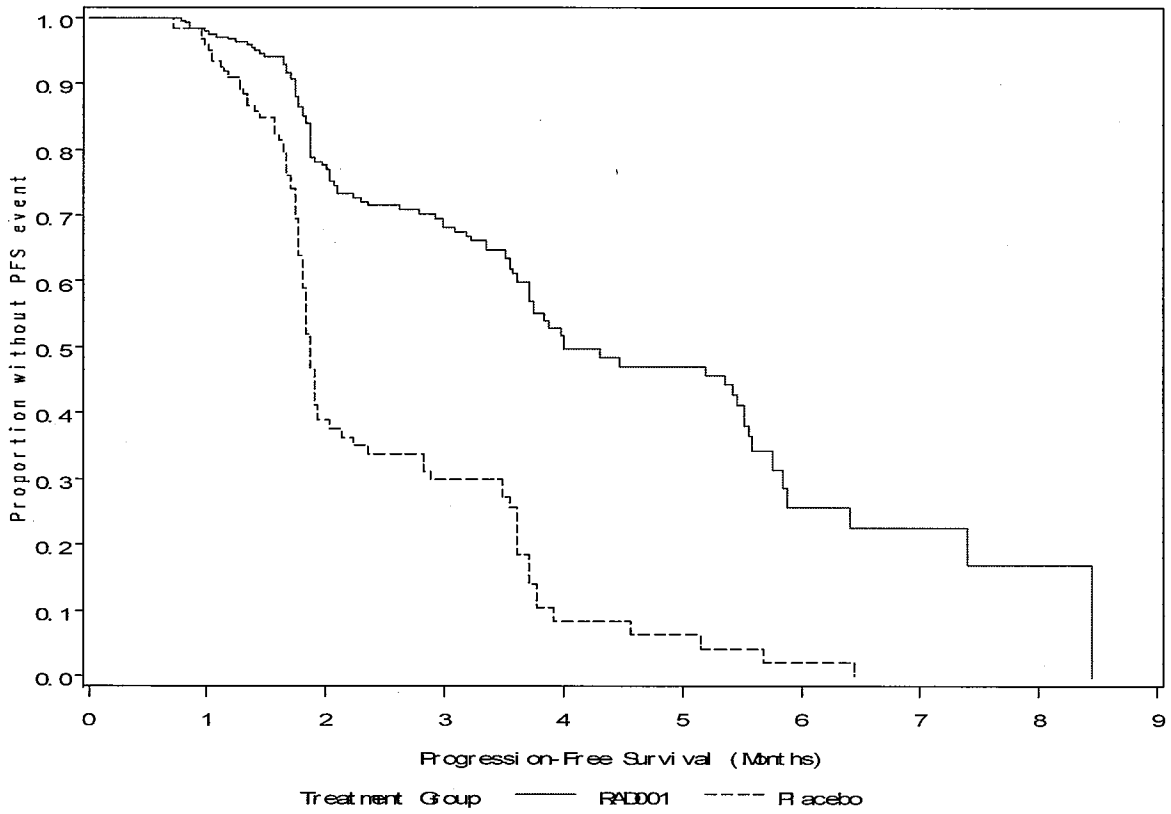


Table 9: Primary Efficacy Analysis of PFS Based on Independent Review (Feb 28, 2008 Cut-off)

Treatment	Number of Patients	Number (%) Failed, Died, Progressed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	277	155 (55.96%) 21 (7.58%) 134 (48.38%)	4.90 (3.98, 5.52)	0.338 (0.262, 0.436)	<0.0001
Placebo	139	111 (79.86%) 8 (5.76%) 103 (74.10%)	1.87 (1.84, 1.94)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis.

Figure 4: Kaplan-Meier Plot of PFS Based on Independent Review (Feb 28, 2008 Cut-off)

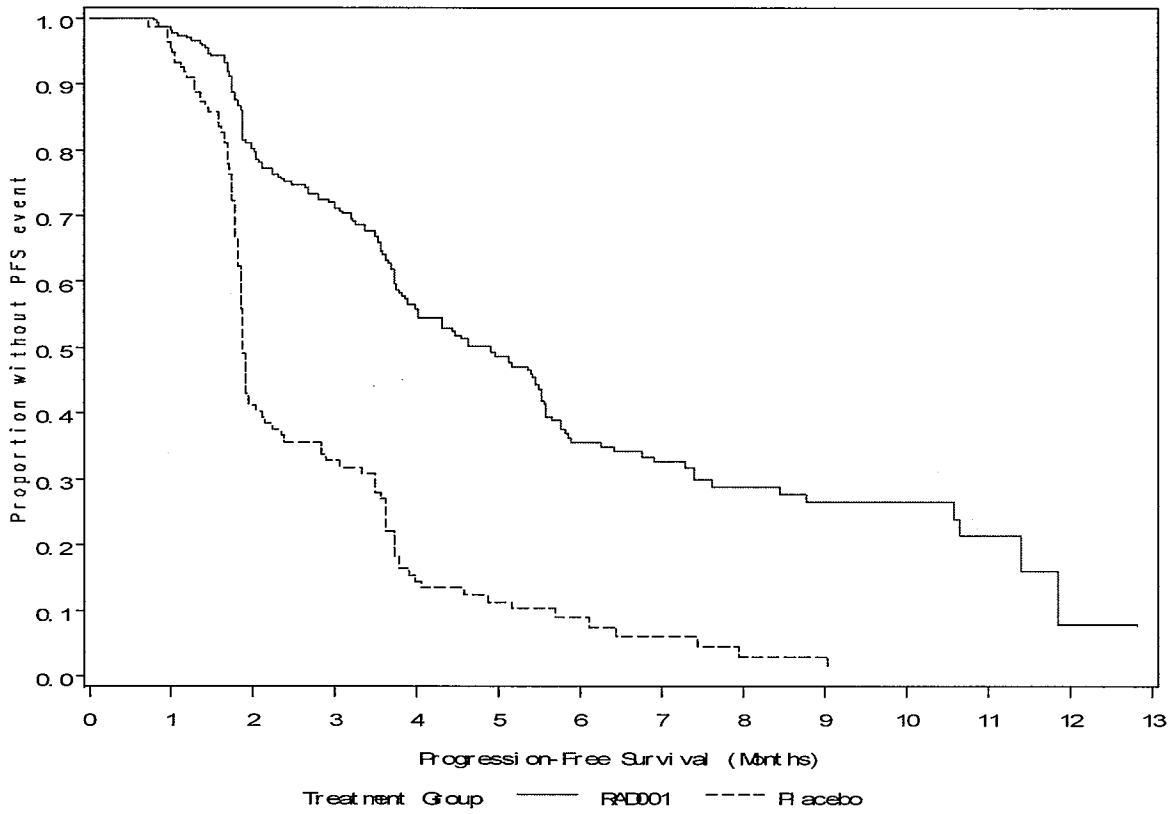


Table 10: Analysis of PFS Based on Investigator Assessment (Oct 15, 2007 Cut-off)

Treatment	Number of Patients	Number (%) Failed, Died, Progressed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	272	111 (40.81%) 14 (5.15%) 97 (35.66%)	4.57 (3.91, 5.52)	0.314 (0.238, 0.414)	<0.0001
Placebo	138	105 (76.09%) 7 (5.07%) 98 (71.01%)	1.84 (1.81, 1.94)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis.

Figure 5: Kaplan-Meier Plot of PFS Based on Investigator Assessment (Oct 15, 2007 Cut-off)

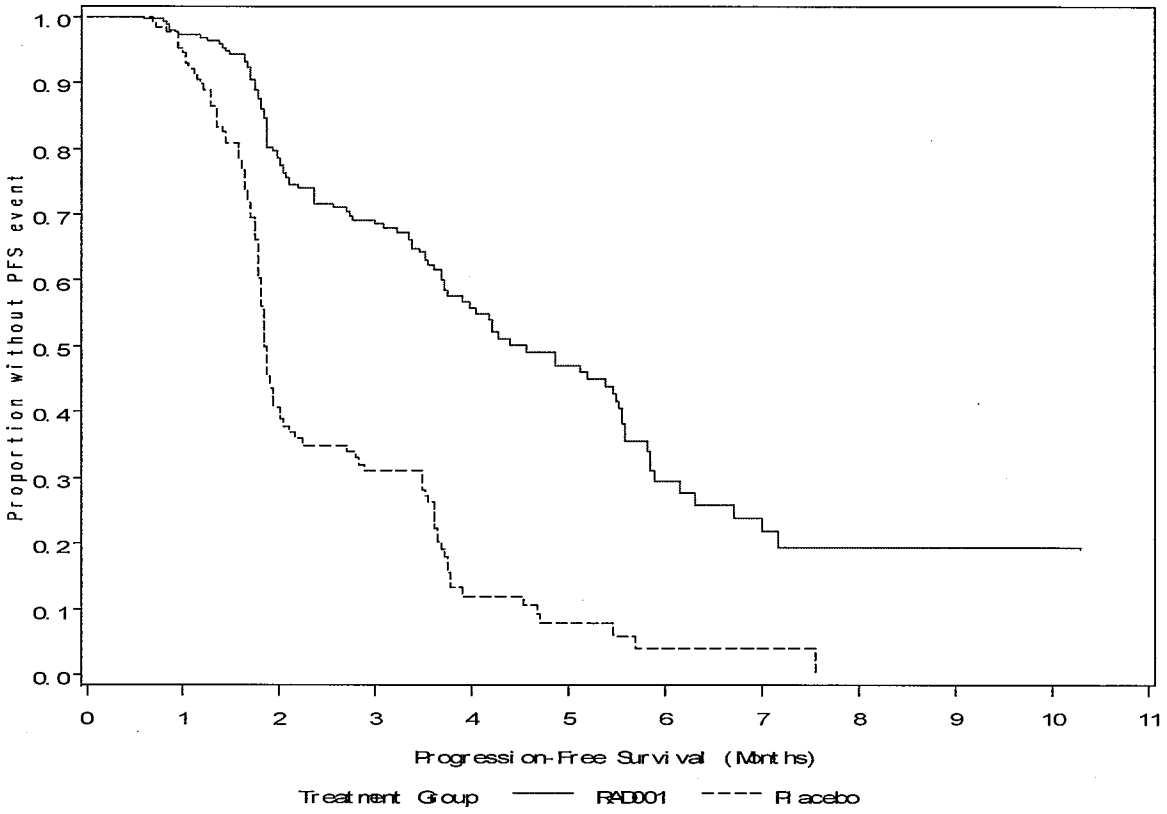


Table 11: Analysis of PFS Based on Investigator Assessment (Feb 28, 2008 Cut-off)

Treatment	Number of Patients	Number (%) Failed, Died, Progressed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	277	170 (61.37%) 18 (6.50%) 152 (54.87%)	5.49 (4.63, 5.82)	0.326 (0.256, 0.414)	<0.0001
Placebo	139	129 (92.81%) 8 (5.76%) 121 (87.05%)	1.87 (1.84, 2.23)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis.

Figure 6: Kaplan-Meier Plot of PFS Based on Investigator Assessment (Feb 28, 2008 Cut-off)

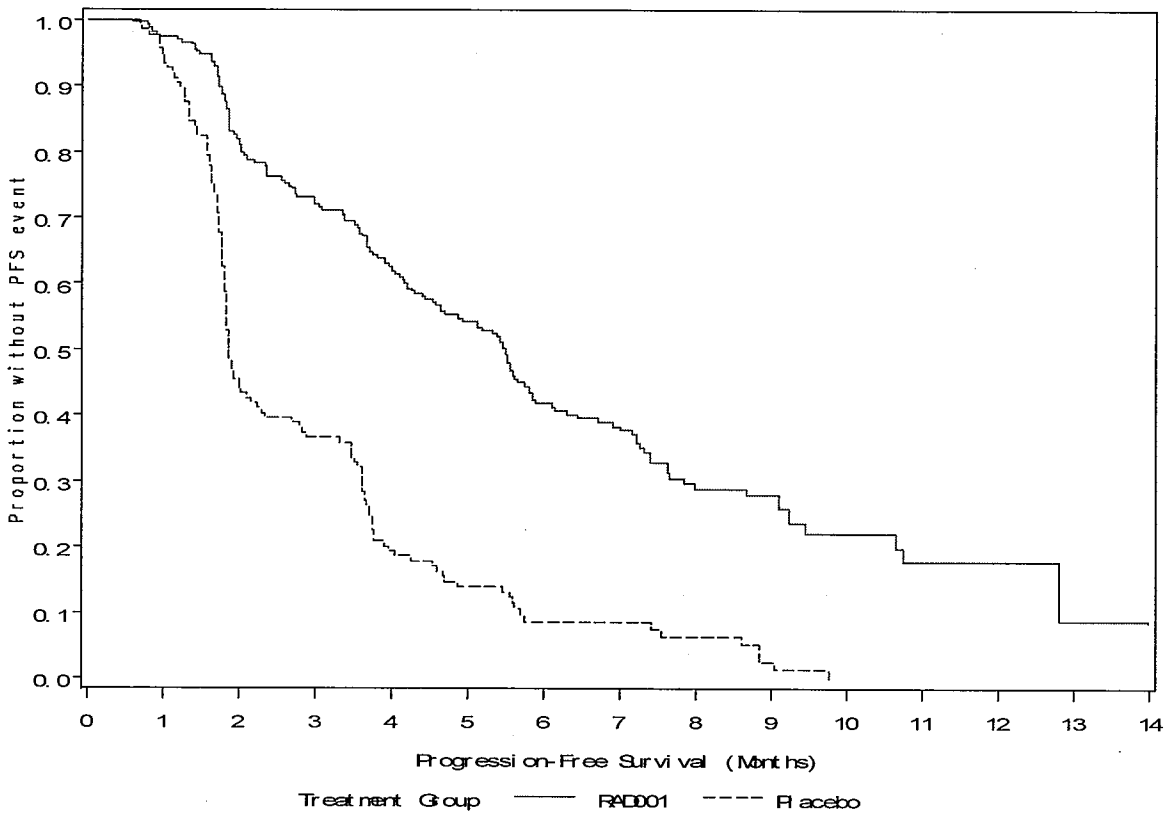


Table 12 shows the number of different types of PFS events and censoring reasons based on independent radiologic review at February 28, 2008 data cut-off.

Table 12: PFS Event Types and Censoring Reasons According to Independent Review (Feb 28, 2008 cut-off)

	RAD001 N=277 (%)	**Placebo N=139 (%)**
PFS events	155 (56.0%)	111 (79.9%)
Death	21 (7.6%)	8 (5.8%)
Disease progression	134 (48.4%)	103 (74.1%)
Censoring	122 (44.0%)	28 (20.1%)
Ongoing without event	54 (19.5%)	4 (2.9%)
Lost to follow-up	2 (0.7%)	0 (0%)
Withdrew consent	8 (2.9%)	0(0%)
Adequate assessment no longer available	20 (7.2%)	4 (2.9%)
New cancer therapy added	34 (12.3%)	20 (14.4%)
Event after >=2 missing tumor assessments	4 (1.4%)	0 (0%)

The disagreement between PFS assessments by investigator and that by independent radiologic review is explored in Table 13, Table 14 and Table 15. Table 13 cross-tabulates disagreement in type of event or censoring (PD, death or censoring) and time of event or censoring (same, time of event or censoring as determined by independent review is later than that by the investigator and time of event or censoring as determined by independent review is earlier than that by the investigator). Table 14 shows disagreement in time of event or censoring and without considering the type of event or censing. Overall summary of disagreements is presented in Table 15.

Table 13: PFS Date and Event Type/Censor Disagreement (Feb 28, 2008 Cut-off)

Event Type		**RAD001 Arm Frequency (%)**				**Placebo Arm Frequency (%)**			
IRC	**INV**	**Same time**	**IRC after INV**	**IRC before INV**	**Total**	**Same time**	**IRC after INV**	**IRC before INV**	**Total**
Censor	PD	33 (11.9)	3 (1.1)	13 (4.7)	49 (17.7)	15 (10.8)	4 (2.9)	2 (1.4)	21 (15.1)
Death	PD	0 (0.0)	10 (3.6)	0 (0.0)	10 (3.6)	0 (0.0)	2 (1.4)	0 (0.0)	2 (1.4)
PD	Censor	9 (3.2)	5 (1.8)	20 (7.2)	34 (12.3)	2 (1.4)	0 (0.0)	1 (0.7)	3 (2.2)
PD	Death	0 (0.0)	0 (0.0)	7 (2.5)	7 (2.5)	0 (0.0)	0 (0.0)	2 (1.4)	2 (1.4)
Censor	Censor	61 (22.0)	4 (1.4)	8 (2.9)	73 (26.4)	6 (4.3)	1 (0.7)	0 (0.0)	7 (5.0)
Death	Death	11 (4.0)	0 (0.0)	0 (0.0)	11 (4.0)	6 (4.3)	0 (0.0)	0 (0.0)	6 (4.3)
PD	PD	50 (18.1)	12 (4.3)	31 (11.2)	93 (33.6)	72 (51.8)	5 (3.6)	21 (15.1)	98 (70.5)
Total		164 (59.2)	34 (12.3)	79 (28.5)	277 (100)	101 (72.7)	12 (8.6)	26 (18.7)	139 (100)

PD: Progressive disease. IRC: As determining by independent radiologic committee. INV: As determined by the investigator.

Table 14: PFS Event/Censor Time Disagreement (Feb 28, 2008 Cut-off)

	RAD001 N=277 Frequency (%)	**Placebo N=139 Frequency (%)**
IRC Event/Censor Time is Before INV's	79 (28.5)	26 (18.7)
IRC Event/Censor Time is After INV's	34 (12.3)	12 (8.6)
More Than 2 Months Difference	60 (21.7)	13 (9.4)
Maximum of IRC Time – INV Time	4.0 Months	3.7 Months
Maximum of INV Time – IRC Time	8.3 Months	6.9 Months

IRC: As determining by independent radiologic committee. INV: As determined by the investigator.

Table 15: Summary of Disagreement between Independent and Investigator assessments of PFS (Feb 28, 2008 Cut-off)

Type of Disagreement	**RAD001 N=277 Frequency (%)**	**Placebo N=139 Frequency (%)**
Disagreement in event type/ censor	100 (36.1)	28 (20.1)
Disagreement in event time	113 (40.8)	38 (27.3)
Any disagreement	155 (56.0)	55 (39.6)
Agreement in both event type/censor and time	122 (44.0)	84 (60.4)

Reviewer's Comments:

1. There was disagreement in type or time of PFS event and censoring between independent and investigator's assessment for a large number of patients (56.0% in RAD001 arm and 39.6% in placebo arm. However the analysis results were similar based on these two assessments. The disagreement between independent and investigator's assessment seems more prominent in the RAD001 arm than in the placebo arm. This indicates a possible bias in investigator's assessment favoring RAD001 in spite of the study being double-blind.
2. Most of the statistical reviewer's analysis results match with those of the applicant. However the hazard ratios and their confidence intervals differ a little. This is due to how the ties are handled in Cox regression. The reviewer used Breslow method whereas the applicant used "discrete" method for handling ties. The applicant's reason for using discrete method is that the data are really discrete in nature and the Breslow method may be biased in presence of large number of tied observations. However, the reviewer compared both the "discrete" and Breslow method and the "exact" method. Although the results are very close for all methods, both "discrete" and Breslow method seem to be

biased, but in different directions. The "discrete" method seems to be biased in favor of the RAD001. See Appendix B for applicant's analysis of PFS.

3. Discrepancy in number of progression events and deaths between patient disposition summary and the efficacy results was discovered during the review of this application. FDA requested the applicant to explain the reasons for this discrepancy. The applicant explained that this apparent discrepancy was because two different sources of information were used for patient disposition and efficacy PFS analysis. The PFS analysis was solely based on radiologic evaluation either by independent central radiology review or by local radiology review according to RECIST whereas death or PD as the primary reason for discontinuation of the study drug was as reported by the investigator on the "End-of-treatment – blinded portion" of the case report form. The patients were followed until progression even after discontinuation of study drug. For primary analysis of PFS, any events occurring after study drug discontinuation were counted as such while patients who were event-free at the data-cut-off date were censored at the last adequate tumor assessment prior to the analysis cut-off and patients who started treatment with a new anticancer therapy before progression were censored at the last adequate tumor assessment prior to the initiation of the new anticancer therapy. These conventions were used for both analyses of PFS by independent review and that by local investigator assessment. Upon FDA's request, the applicant submitted the analyses of PFS after censoring a patient at the last adequate tumor assessment before study drug discontinuation if patient did not have any event or otherwise censored before the study drug discontinuation. The reviewer has confirmed the results. The efficacy conclusions did not change in that analysis. Although the hazard ratios in that analysis were slightly smaller than those in the primary analysis results, the confidence intervals of the hazard ratio were wider. In addition the reviewer has calculated the time from discontinuation of study drug to the PFS event or censoring time for the patients whose PFS event or censoring occurs after study drug discontinuation. Summaries of these calculations based on February 28, 2008 data cut-off date are presented in Table 16, Table 17, Table 18 and Table 19.

Table 16: Summary of time (in days) from study drug discontinuation to PFS event or censoring based on independent radiology review for the patients who have PFS event or are censored after study drug discontinuation

	RAD001	**Placebo**	**All**
N	66	36	102
Mean	28.6	10.8	22.3
First quartile	4	1	1
Median	21	6	10
Third quartile	39	9.5	33
Maximum	126	71	126
Number of patients with more than 28 days	23	4	27

Table 17: Summary of time (in days) from study drug discontinuation to PFS event based on independent radiology review for the patients who have PFS event after study drug discontinuation

	RAD001	Placebo	All
N	50	32	82
Mean	32.6	11.6	24.4
First quartile	6	1	1
Median	23	6	11.5
Third quartile	49	10	37
Maximum	126	71	126
Number of patients with more than 28 days	21	4	25

Table 18: Summary of time (in days) from study drug discontinuation to PFS event or censoring based on local radiology review for the patients who have PFS event or are censored after study drug discontinuation

	RAD001	Placebo	All
N	69	35	104
Mean	20.1	8.9	16.3
First quartile	1	1	1
Median	12	1	7
Third quartile	27	8	23
Maximum	105	55	105
Number of patients with more than 28 days	16	4	20

Table 19: Summary of time (in days) from study drug discontinuation to PFS event based on local radiology review for the patients who have PFS event after study drug discontinuation

	RAD001	Placebo	All
N	60	35	95
Mean	19.4	8.9	15.6
First quartile	1	1	1
Median	11	1	7
Third quartile	25.5	8	23
Maximum	105	55	105
Number of patients with more than 28 days	14	4	18

3.1.7.5. Secondary Efficacy Analyses

The analyses of secondary endpoint OS based on the data with cut-off dates October 15, 2007 and February 28, 2008 are presented in Table 20 and Table 21, respectively. The corresponding Kaplan-Meier plots are given in Figure 7 and Figure 8. RAD001 did not show statistically significant improvement in OS over placebo (log-rank test, nominal one-sided p-value 0.234 and 0.138 for the two cut-off dates adjusted for MSKCC risk group and not adjusted for interim analyses). The hazard ratios and their 95% confidence intervals for OS based on the two cut-off dates (Oct 15, 2007 and February 28, 2008) were 0.831 [95% CI: (0.505, 1.370)] and 0.821 [95% CI: (0.575, 1.171)], respectively. The tabulations of best overall response and analyses of ORR as determined by the independent radiologic review are presented in Table 22 and Table 23 respectively for the two cut-off dates. The overall response rate was extremely low in both arms (<2% in RAD001 vs. 0% in placebo) and was not statistically significant (Cochran-Mantel-Haenszel test, p-value 0.216 and 0.113, respectively for the two cut-off dates). For the updated data, the odds ratio was 2.21 with a 95% confidence interval (0.37, 13.34). Other secondary endpoints were time to definitive deterioration in Functional Assessment of Cancer Therapy Kidney Symptom Index - Disease Related Symptoms (FKSI-DRS) and the physical functioning (PF) and quality of life (QL) scales of European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30. Since neither OS nor ORR was statistically significant, other secondary endpoints were not formally tested.

Table 20: Analysis of OS (Oct 15, 2007 Cut-off)

Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	272	42 (15.44%)	NE (7.23, NE)	0.831	0.234
Placebo	138	26 (18.84%)	8.80 (7.92, NE)	(0.505, 1.370)	

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis. NE: Not estimable.

Figure 7: Kaplan-Meier Plot of OS (Oct 15, 2007 Cut-off)

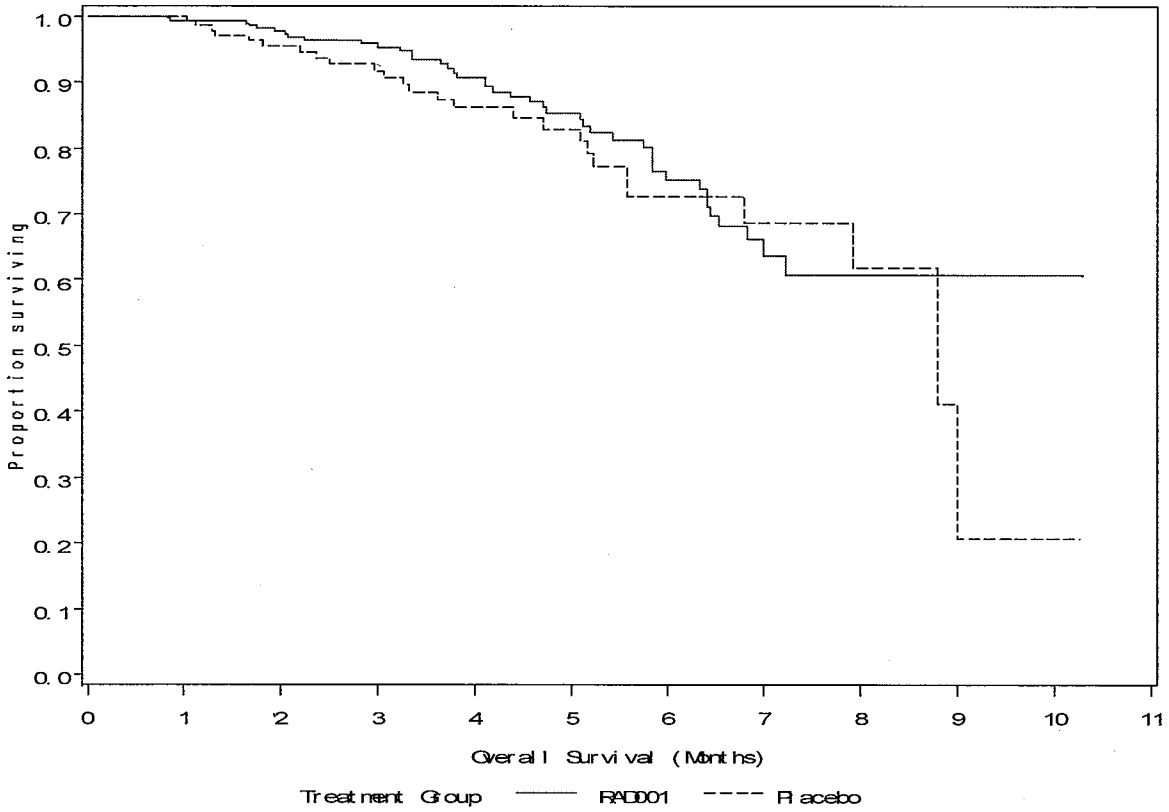


Table 21: Analysis of OS (Feb 28, 2008 Cut-off)

Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)	P-value[3]
RAD001	277	85 (30.69%)	NE (11.43, NE)	0.821 (0.575, 1.171)	0.138
Placebo	139	48 (34.53%)	13.01 (10.09, NE)		

[1]: Kaplan-Meier estimate. [2]: Based on Cox model stratified by MSKCC risk group. [3]: Based on one-sided log-rank test adjusted for MSKCC risk group stratification factor, not adjusted for interim analysis. NE: Not estimable.

Figure 8: Kaplan-Meier Plot of OS (Feb 28, 2008 Cut-off)

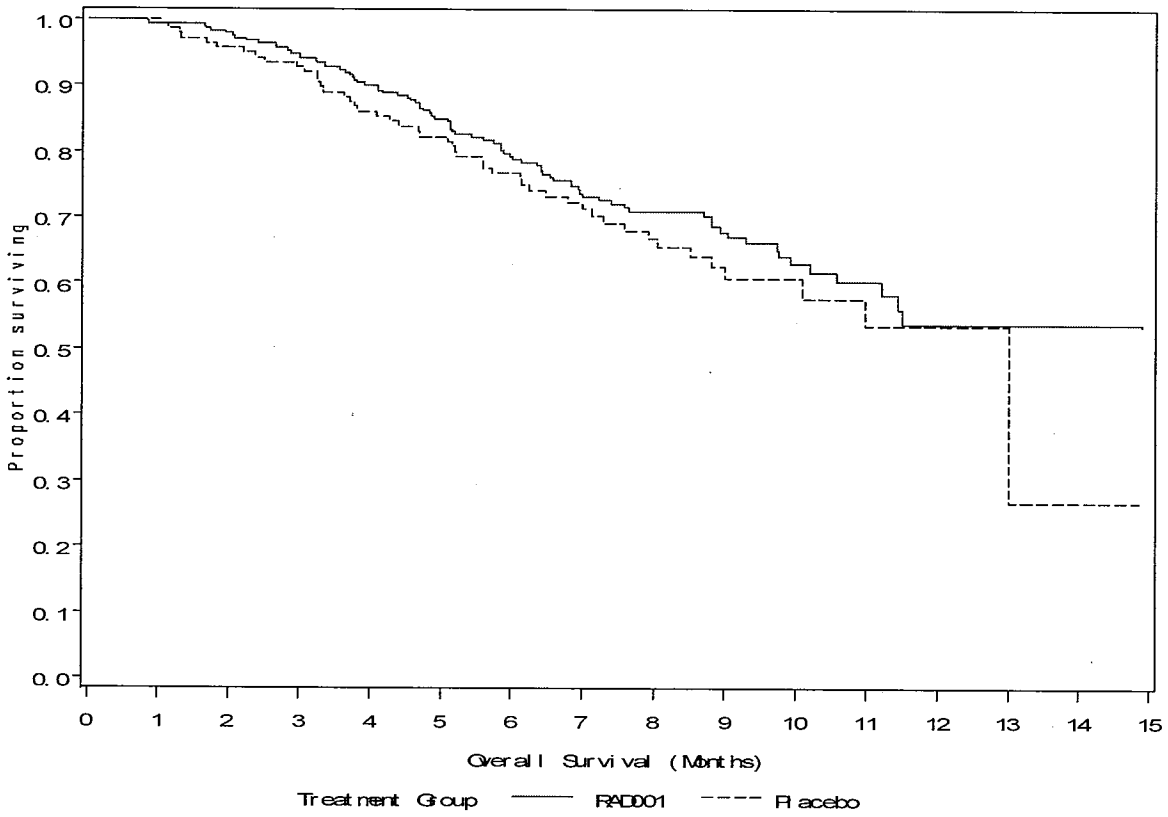


Table 22: Best Overall Tumor Response Based on Independent Review (Oct 15, 2007 Cut-off)

	Randomization Group		
	RAD001 (N=272)	**Placebo (N=138)**	**All (N=410)**
Complete Response	0	0	0
Partial Response	3 (1.10%)	0	3 (0.73%)
Stable Disease	171 (62.87%)	44 (31.88%)	215 (52.44%)
Progressive Disease	53 (19.49%)	63 (45.65%)	116 (28.29%)
Unknown	45 (16.54%)	31 (22.46%)	76 (18.54%)
Overall Response Rate (95% CI)	1.10% (0.23%, 3.19%)	0%	0.73%
P-value[1]	0.216		

[1]: Based on Cochran-Mantel Haenszel test stratified by MSKCC risk group.

Table 23: Best Overall Tumor Response Based on Independent Review (Feb 28, 2008 Cut-off)

	Randomization Group		
	RAD001 (N=277)	**Placebo (N=139)**	**All (N=416)**
Complete Response	0	0	0
Partial Response	5 (1.81%)	0	5 (1.20%)
Stable Disease	185 (66.79%)	45 (32.37%)	230 (55.29%)
Progressive Disease	57 (20.58%)	74 (53.24%)	131 (31.49%)
Unknown	30 (10.83%)	20 (14.39%)	50 (12.02%)
Overall Response Rate (95% CI)	1.81% (0.59%, 4.16%)	0%	1.20%
P-value[1]	0.113		

[1]: Based on Cochran-Mantel Haenszel test stratified by MSKCC risk group.

Reviewer's Comments:

1. The statistical reviewer's hazard ratio and the confidence intervals in the OS analyses do not match with those of the applicant because of use of different methods for handling ties as stated before. See Appendix B for applicant's analysis of OS.
2. Since 109 (78.4%) of 139 patients receiving placebo were crossed over to RAD001 on disease progression, it is difficult to evaluate the true effect of RAD001 on survival.

3.2. Evaluation of Safety

For safety evaluation, please refer to the clinical review of this application.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1. Gender, Race and Age

Efficacy by gender was analyzed by exploratory analysis of PFS and is presented in Table 24. Efficacy by age group (<65 years, ≥65 years) was also analyzed by exploratory analysis of PFS and is presented in Table 25. Exploratory analyses of OS by gender and by age group are also performed and presented in Table 26 and Table 27, respectively. No analysis has been performed by race because majority of the subjects were Caucasian (88.2%) and the second largest racial group Asian constitutes only 6.5% of the subjects. All analyses in this section are based on updated data with cut-off date February 28, 2008.

Table 24: Exploratory Analysis of PFS Based on Independent Review by Gender

Gender	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
Female	RAD001	61	35 (57.38%)	5.13 (3.35, 5.88)	0.398 (0.228, 0.693)
	Placebo	33	26 (78.79%)	1.91 (1.71, 3.61)	
Male	RAD001	216	120 (55.56%)	4.90 (4.01, 5.52)	0.336 (0.251, 0.450)
	Placebo	106	85 (80.19%)	1.87 (1.84, 1.94)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria.

Table 25: Exploratory Analysis of PFS Based on Independent Review by Age Group

Age Group	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
<65 Years	RAD001	165	93 (56.36%)	4.30 (3.71, 5.49)	0.362 (0.264, 0.495)
	Placebo	98	79 (80.61%)	1.87 (1.84, 1.91)	
≥65 Years	RAD001	112	62 (55.36%)	5.36 (3.98, 5.88)	0.340 (0.217, 0.532)
	Placebo	41	32 (78.05%)	2.23 (1.81, 3.48)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria.

Table 26: Exploratory Survival Analysis by Gender

Gender	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
Female	RAD001	61	23 (37.70%)	11.43 (6.97, NE)	1.170 (0.521, 2.625)
	Placebo	33	10 (30.30%)	13.01 (NE, NE)	
Male	RAD001	216	62 (28.70%)	NE (11.20, NE)	0.725 (0.482, 1.089)
	Placebo	106	38 (35.85%)	10.97 (8.80, NE)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria. NE: Not estimable.

Table 27: Exploratory Survival Analysis by Age Group

Age Group	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio RAD001/Placebo (95% CI)
<65 Years	RAD001	165	58 (35.15%)	NE (10.58, NE)	0.923 (0.606, 1.403)
	Placebo	98	36 (36.73%)	10.97 (9.00, 13.01)	
≥ 65 Years	RAD001	112	27 (24.11%)	NE (11.43, NE)	0.717 (0.362, 1.420)
	Placebo	41	12 (29.27%)	NE (8.80, NE)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria. NE: Not estimable.

Reviewer's Comments:

1. The PFS results are similar across the gender and age groups.
2. In the female subjects the hazard ratio of RAD001 to placebo for OS is greater than 1 which appears to indicate that RAD001 may not be beneficial for female patients in terms of OS. However, because of a small number of female patients, the confidence interval for the hazard ratio is wide and a conclusion cannot be drawn based on this.
3. The OS results across the age group are similar, although RAD001 appears to be more effective in the subjects ≥ 65 years of age which may also be due to slight imbalance in randomization across the age groups.

4.2. Other Special/Subgroup Populations

Exploratory analyses of PFS and OS by geographic regions are presented in Table 28 and Table 29, respectively. The stratification factors at randomization were MSKCC risk criteria (favorable vs. intermediate vs. poor) and number of prior VEGFr-TKI therapies (1 vs. 2). Exploratory analyses of PFS by MSKCC risk criteria and prior VEGFr-TKI therapy (sorafenib, sunitinib, both) are presented in Table 30 and Table 31, respectively. All analyses are based on February 28, 2008 cut-off date (updated data).

Table 28: Exploratory Analysis of PFS Based on Independent Review by Geographic Region

Geographic Region	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
USA and Canada	RAD001	77	43 (55.84%)	4.63 (3.71, 5.59)	0.308 (0.193, 0.492)
	Placebo	53	41 (77.36%)	1.87 (1.77, 2.10)	
Europe	RAD001	180	104 (57.78%)	4.44 (3.71, 5.52)	0.422 (0.299, 0.594)
	Placebo	71	59 (83.10%)	1.91 (1.81, 2.83)	
Australia and Japan	RAD001	20	8 (40.00%)	10.58 (4.90, NE)	0.184 (0.060, 0.568)
	Placebo	15	11 (73.33%)	1.87 (1.84, 3.61)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria. NE: Not estimable.

Table 29: Exploratory Survival Analysis by Geographic Region

Geographic Region	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
USA and Canada	RAD001	77	26 (33.77%)	11.43 (9.92, NE)	1.023 (0.553, 1.894)
	Placebo	53	19 (35.85%)	13.01 (9.00, NE)	
Europe	RAD001	180	56 (31.11%)	NE (11.20, NE)	0.803 (0.505, 1.277)
	Placebo	71	27 (38.03%)	10.97 (7.59, NE)	
Australia and Japan	RAD001	20	3 (15.00%)	NE (10.58, NE)	1.502 (0.147,15.385)
	Placebo	15	2 (13.33%)	NE (7.92, NE)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria. NE: Not estimable.

Table 30: Exploratory Analysis of PFS Based on Independent Review by MSKCC Risk Group

MSKCC Risk Group	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
Favorable	RAD001	81	39 (48.15%)	5.75 (3.98, 7.39)	0.317 (0.196, 0.511)
	Placebo	39	33 (84.62%)	1.94 (1.87, 2.83)	
Intermediate	RAD001	156	90 (57.69%)	4.53 (3.78, 5.49)	0.324 (0.231, 0.453)
	Placebo	79	61 (77.22%)	1.84 (1.81, 1.94)	
Poor	RAD001	40	26 (65.00%)	3.55 (1.87, 4.63)	0.448 (0.232, 0.867)
	Placebo	21	17 (80.95%)	1.84 (1.77, 3.61)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model.

Table 31: Exploratory Analysis of PFS Based on Independent Review by Prior VEGFr-TKI Therapy

Prior VEGFr-TKI Therapy	Treatment	Number of Patients	Number (%) Failed	Median in Months[1] (95% CI)	Hazard Ratio[2] RAD001/Placebo (95% CI)
Sorafenib	RAD001	81	37 (45.68%)	5.88 (4.90, 11.40)	0.264 (0.163, 0.430)
	Placebo	43	36 (83.72%)	2.83 (1.91, 3.61)	
Sunitinib	RAD001	124	69 (55.65%)	3.88 (3.55, 5.55)	0.373 (0.253, 0.551)
	Placebo	60	47 (78.33%)	1.84 (1.77, 1.91)	
Both	RAD001	72	49 (68.06%)	4.01 (3.55, 5.36)	0.321 (0.192, 0.537)
	Placebo	36	28 (77.78%)	1.84 (1.77, 2.04)	

[1]: Kaplan-Meier estimate. [2]: Proportional hazards model stratified by MSKCC risk criteria.

Reviewer's Comments:

1. RAD001 appears to be most effective in the European patients in terms of OS but least effective in that population in terms of PFS.
2. RAD001 appears to be ineffective in non-European regions in terms of OS. However, due to small number of patients and OS events in non-European population, no conclusions can be made.
3. RAD001 appears to be less effective in the poor risk patients than in the favorable and intermediate risk patients in terms of PFS.
4. Hazard ratios of RAD001 to placebo are similar across prior VEGFr-TKI therapies.

5. SUMMARY AND CONCLUSIONS

This application, supported by Study C2240, seeks the approval of Afinitor (everolimus or RAD001) for the indication of advanced renal cell carcinoma (RCC). This double-blind, randomized, stratified, multicenter, international, phase III trial compares RAD001 plus best supportive care (BSC) to placebo plus BSC in the metastatic RCC patients whose disease has progressed despite prior therapy with sunitinib, sorafenib or both sequentially. This study randomized a total of 416 patients, 277 to RAD001 and 139 to placebo. The primary efficacy endpoint was progression-free survival (PFS). The secondary efficacy endpoints were overall survival (OS), overall response rate (ORR), and time to definitive deterioration in Functional Assessment of Cancer Therapy Kidney Symptom Index - Disease Related Symptoms (FKSI-DRS) and the physical functioning (PF) and quality of life (QL) scales of European Organization for Research and Treatment of Cancer (EORTC) QLQ-C30. The study stopped at a planned interim analysis conducted on February 25, 2008 for efficacy based on data cut-off date of October 15, 2007 and the study patients were unblinded on February 28, 2008. The RAD001 arm showed statistically significant improvement over placebo arm with respect to PFS as determined by independent radiologic review (log-rank test, one sided p-value < 0.0001 adjusted for MSKCC risk group and hazard ratio 0.338) in the ITT population based on February 28, 2008 data cut-off date. The main secondary endpoints OS and ORR were not statistically different between RAD001 and placebo arms (log-rank test, p-value=0.138 for one-sided and adjusted for MSKCC risk group for OS and Cochran-Mantel-Haenszel test stratified by MSKCC risk group, p-value=0.113 for ORR).

5.1. Statistical Issues and Collective Evidence

1. The quality of the datasets and the whole submission is poor. There were many mistakes and inconsistencies in the dataset that were revealed only after requests for clarification by the reviewer. Many documents were incomplete or not present at all. Some of them were submitted after the reviewers identified and requested them. Because of the poor quality of the datasets, inaccuracy of the results is a concern. For details of FDA's correspondence with the applicant, please refer to Appendix A.
2. Consideration of PFS as the primary endpoint for demonstration of efficacy for approval of drug products is based on the magnitude of the effect and the risk-benefit profile of the drug product. Because documentation of PFS assessments are often based on both subjective and objective criteria and these assessments depend on frequency, accuracy, reproducibility and completeness of tumor assessments, it is important that the observed magnitude of effect is robust. An interim PFS analysis may not provide an accurate or reproducible estimate of the treatment effect size due to inadequate follow-up, missing assessments, disagreements between radiological reviewers and/or disagreements between investigator and independent assessments. Stopping a trial based on interim PFS results which may not be verifiable after adjudication can be problematic and the trial results, in particular, may not be interpretable if the treatment in the control group was changed based on the interim results. In this application there was a substantial amount of disagreement between the independent and investigator's assessments of PFS. In

addition, the design allowed the patients in placebo arm to crossover to RAD001 arm upon progression as assessed by the investigator. Although the effect of RAD001 on PFS is large, the robustness of this effect is in question. Nevertheless, since the updated data had approximately 91.7% of the events that were required for the final analysis, the lack of robustness of the PFS results is expected to be of a lesser degree.

3. There was disagreement in type or time of PFS event and censoring between independent and investigator's assessment for a large number of patients (56.0% in RAD001 arm and 39.6% in placebo arm. However the analysis results were similar based on these two assessments. The disagreement between independent and investigator's assessment seems more prominent in the RAD001 arm than in the placebo arm. This indicates a possible bias in investigator's assessment favoring RAD001 in spite of the study being double-blind.
4. Since the patients receiving placebo were crossed over to RAD001 on disease progression as assessed by the investigator and also when the study is terminated due to superior efficacy finding at the second interim analysis, it is difficult to evaluate the true effect of RAD001 on survival.

5.2. Conclusions and Recommendations

The applicant has submitted results from one phase III, randomized, double-blind, comparative clinical trial (Study C2240) comparing Afinitor (everolimus or RAD001) plus best supportive care (BSC) and placebo plus BSC in patients with advanced renal cell carcinoma (RCC) who were previously treated with sunitinib, sorafenib or both sequentially. The study showed benefit of RAD001 over placebo in terms of progression-free survival (PFS) as determined by independent radiologic review in this patient population based on the data from a planned interim analysis. However, the overall survival, a secondary endpoint, was not improved with RAD001 with approximately 32% overall deaths, but a trend favoring RAD001 was observed. RAD001 also did not show statistically significant superiority over placebo in terms of overall response rate (another secondary endpoint) as determined by independent radiologic review. The data and statistical results provide adequate evidence to support the claims about PFS proposed in the NDA.

APPENDIX A: STATISTICAL DEFICIENCIES AND INFORMATION REQUESTS

This application was incomplete and had many errors at the initial submission. Many deficiencies have been discovered during the review and information requests have been sent to the applicant. There were 40 submissions including the initial submission in a single review cycle for this NDA till March 3, 2009. Many of the submissions are responses to information requests from the reviewers from different disciplines. Following is a list of information requests and submissions related to statistical review.

Table 32: Problems Found and Information Requests Made to the Applicant during Review

Problem found / Information request sent	Information request sent on	Applicant's response received on	Response satisfactory?
SAS transport dataset containing the user-defined formats had errors	July 23, 2008	July 29, 2008	Yes
Submission (dated August 26, 2008) of main efficacy results based on February 28, 2008 cut-off dare did not include any data	September 22, 2008	September 30, 2008	Yes
Detailed efficacy update (based on February 28, 2008 cut-off date) similar to the original clinical study report was not submitted (clinical request)	September 22, 2008	September 29, 2008	No
Some PFS-related variables could not be identified	September 23, 2008	September 29, 2008	Yes
Multiple variables that were likely to contain the same information existed in the datasets. Those variables did not exactly match. It was not clear which variables were used for efficacy	September 24, 2008 (T-con)	October 14, 2008	Yes
Details of discrepancies between central and local radiology assessments with comments for possible explanation were requested	September 24, 2008 (T-con)	October 20, 2008	Yes
Details of missing tumor assessments based on both central and local reviewers were requested	September 24, 2008 (T-con)	October 24, 2008	Yes
There were discrepancies between raw and derived datasets containing antineoplastic therapies after discontinuation of study drugs based on October 15, 2007 cut-off date	September 24, 2008 (T-con)	November 11, 2008	Yes

There were discrepancies between datasets containing antineoplastic therapies after discontinuation of study drugs based on October 15, 2007 and February 28, 2008 cut-off dates	January 29, 2009	January 30, 2009	No
Applicant's clarification for the above discrepancies between datasets containing antineoplastic therapies after discontinuation of study drugs based on October 15, 2007 and February 28, 2008 cut-off dates was not satisfactory	February 2, 2009	February 4, 2009	Partially
Subgroup analyses by gender, age group, MSKCC risk score and Prior VEGFr-TKI therapy were not submitted for the data with cut-off date February 28, 2008	February 17, 2009	February 18, 2009	Yes
Number of patients with progressive disease or death did not match between safety update table and efficacy update based on cut-off date February 28, 2008 (clinical request)	February 17, 2009, February 19, 2009 and	February 18, 2009, February 23, 2009 and	No
The summary of prior VEGFr-TKI therapy based on submitted data did not match with that in the clinical study report (clinical request)	February 20, 2009	February 23, 2009	No
Independent data monitoring (IDMC) charter, IDMC meeting minutes for the first interim analysis were not submitted	February 23, 2009	February 25, 2009	Yes
IDMC meeting minutes for the second interim analysis could not be located in the submission	February 23, 2009	February 25, 2009	Yes
A table containing the PFS censoring rules in protocol post-text supplement that the statistical analysis plan referred to was not present	February 23, 2009	February 25, 2009	Yes
Applicant's explanation for mismatch of number of PD and deaths between patient disposition and efficacy results was not clear and clarification was requested.	March 6, 2009	March 9 (meeting) and March 10, 2008	Yes
Analysis of PFS after censoring PFS at the last assessment before study drug discontinuation for patients without event or other causes of censoring at the time of study drug discontinuation was requested (the primary analysis of PFS did not	March 6, 2009	March 10, 2009	Yes

censor patients at the time of study drug discontinuation)			

There were several other problems that have not been listed above. Some of these problems are following.

- Time format between variables is inconsistent; some are recorded as number of seconds, some as number of days. These were not indicated in the define.pdf file.
- Codes for some coded variables were neither explained in define.pdf nor were they recorded using user-defined formats.
- Some new datasets with new variables have been created for the data with cut-off date February 28, 2008. The old variables that were present for earlier cut-off date remained in the data and sometimes did not match with the new variables leading to inconsistency.
- There are other extraneous variables that do not serve any purpose. Sometimes they were for the applicant's internal use and still have been kept in the data.
- Errata to the clinical study report for the pivotal study were submitted on December 5, 2008. The cover letter of that submission did not mention it at all.
- The applicant's responses to information requests were sometimes incorrect and misleading.

APPENDIX B: APPLICANT'S RESULTS FOR PFS AND OS

The applicant's analyses of PFS as assessed by independent radiologic review and by investigator based on October 15, 2007 data cut-off date are presented in Table 33 and Table 34, respectively and those analyses based on February 28, 2008 cut-off date are presented in Table 35 and Table 36, respectively. The applicant analyses of OS for October 15, 2007 and February 28, 2008 are presented in Table 37 and Table 38, respectively.

Table 33: Applicant's Analysis of PFS Based on Central Radiology Review Using the Kaplan- Meier Method – ITT Population (October 15, 2007 Data Cut-off Date)

	RAD001 10mg/day (plus BSC) N=272	Placebo (plus BSC) N=138	p-value [1]	Hazard ratio [2] [95% CI] RAD001 / Placebo
No. of PFS events n (%)	101 (37.1)	90 (65.2)	<0.001	0.30 [0.22,0.40]
Progression	85 (31.3)	82 (59.4)		
Death	16 (5.9)	8 (5.8)		
No. censored	171 (62.9)	48 (34.8)		
Kaplan-Meier estimates [95% CI] at:				
4 months	51.9 [43.7;60.1]	8.4 [1.5;15.2]		
6 months	25.7 [14.4;36.9]	2.1 [NA;6.0]		
25th percentile for PFS [95% CI] (months)	2.07 [1.87;2.99]	1.71 [1.61;1.77]		
Median PFS [95% CI] (months)	4.01 [3.71;5.52]	1.87 [1.81;1.94]		
75th percentile for PFS [95% CI] (months)	6.41 [5.59;8.44]	3.61 [2.37;3.71]		

[1] p-value is obtained from the stratified Log-Rank test

[2] Hazard ratio is obtained from stratified Cox model

Table 34: Applicant's Analysis of PFS Based on Investigator Using the Kaplan- Meier Method – ITT Population (October 15, 2007 Data Cut-off Date)

	RAD001 10mg/day (plus BSC) N=272	Placebo (plus BSC) N=138	p-value [1]	Hazard ratio [2] [95% CI] RAD001 / Placebo
No. of PFS events n (%)	111 (40.8)	105 (76.1)	<0.001	0.31 [0.23,0.41]
Progression	97 (35.7)	98 (71.0)		
Death	14 (5.1)	7 (5.1)		
No. censored	161 (59.2)	33 (23.9)		
Kaplan-Meier estimates [95% CI] at:				
4 months	55.9 [48.3;63.4]	11.9 [5.4;18.5]		
6 months	29.6 [20.4;38.7]	4.0 [NA;8.8]		
25th percentile for PFS [95% CI] (months)	2.10 [1.87;3.09]	1.64 [1.45;1.74]		
Median PFS [95% CI] (months)	4.57 [3.91;5.52]	1.84 [1.81;1.94]		
75th percentile for PFS [95% CI] (months)	6.70 [5.85;NA]	3.61 [2.79;3.75]		

[1] p-value is obtained from the stratified Log-Rank test

[2] Hazard ratio is obtained from stratified Cox model

Table 35: Applicant's Analysis of PFS Based on Central Radiology Review – ITT Population (February 28, 2008 Data Cut-off Date)

Primary endpoint	Number (%) of patients with PFS event		Comparison between groups		
	Everolimus N=277	Placebo N=139	Hazard ratio [a]	95% CI [a]	p-value [b]
Primary analysis: independent central review	155 (56.0)	111 (79.9)	0.33	0.25 to 0.43	<0.001

[a] Cox model; [b] One-sided stratified log-rank test

Table 36: Applicant's Analysis of PFS Based on Investigator – ITT Population (February 28, 2008 Data Cut-off Date)

Primary endpoint	Number (%) of patients with PFS event		Comparison between groups		
	Everolimus N=277	Placebo N=139	Hazard ratio [a]	95% CI [a]	p-value [b]
Supportive analysis: investigator assessment	170 (61.4)	129 (92.8)	0.32	0.25 to 0.41	<0.001

[a] Cox model; [b] One-sided stratified log-rank test

Table 37: Applicant's Analysis of OS Using the Kaplan- Meier Method – ITT Population (October 15, 2007 Data Cut-off Date)

	RAD001 10mg/day (plus BSC) N=272	Placebo (plus BSC) N=138	p-value [1]
No. of events - n (%)	42 (15.4)	26 (18.8)	0.233
No. of censored	230 (84.6)	112 (81.2)	
Kaplan-Meier estimates [95% CI] at:			
4 months	90.7 [86.5;94.9]	86.1 [79.4;92.8]	
6 months	75.3 [67.4;83.2]	72.8 [62.0;83.6]	
25th percentile OS [95% CI] (months)	6.34 [5.42;7.00]	5.59 [5.09;9.00]	
Median OS [95% CI] (months)	NA [7.23;NA]	8.80 [7.92;NA]	
75th percentile OS [95% CI] (months)	NA [NA;NA]	9.00 [8.80;NA]	

[1] p-value is obtained from the stratified Log-Rank test

Table 38: Applicant's Analysis of OS – ITT Population (February 28, 2008 Data Cut-off Date)

Overall survival	Number (%) of deaths		Comparison between groups		
	Everolimus N=277	Placebo N=139	Hazard ratio [a]	95% CI [a]	p-value [b]
Primary analysis	85 (30.7)	48 (34.5)	0.82	0.57 to 1.17	0.137

[a] Cox model; [b] One-sided stratified log-rank test

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Somesh Chattopadhyay, Ph.D.
Date: March 18, 2009

Concurring Reviewer(s): Shenghui Tang, Ph.D., Acting Team Leader
Rajeshwari Sridhara, Ph.D., Deputy Director

Statistical Team Leader: Shenghui Tang, Ph.D.

Biometrics Division Director: Aloka Chakravarty, Ph.D.

cc:
HFD-150/Ms. Christy Cottrell
HFD-150/Dr. Qin Ryan
HFD-150/Dr. Robert Justice
HFD-711/Dr. Somesh Chattopadhyay
HFD-711/Dr. Shenghui Tang
HFD-711/Dr. Rajeshwari Sridhara
HFD-711/Dr. Aloka Chakravarty
HFD-700/Ms. Lilian Patrician

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Somesh Chattopadhyay
3/18/2009 10:03:45 AM
BIOMETRICS

Shenghui Tang
3/18/2009 10:11:47 AM
BIOMETRICS

Rajeshwari Sridhara
3/18/2009 10:47:16 AM
BIOMETRICS