CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-560s000

STATISTICAL REVIEW(S)

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-560 / 0000
Drug Name:	everolimus Tablets, 0.25 mg, 0.5 mg, 0.75 mg, 1.0 mg (formerly SDZ, RAD, Certican)
Indication(s):	Prophylaxis of organ rejection in allogeneic kidney recipients
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submitted June 30, 2009
Review Priority:	Standard (6-month)
Biometrics Division:	Division of Biometrics IV
Statistical Reviewer:	Xiao Ding, Ph.D.
Concurring Reviewers:	LaRee Tracy, Ph.D.
	Karen Higgins, Sc.D.
	Mohammad Huque, Ph.D.
Medical Division:	Division of Special Pathogen and Transplant Products
Clinical Team:	Medical Officer: Ergun Velidedeoglu, M.D.
	Medical Team Leader: Joette Meyer, Pharm. D.
Project Manager:	Jacquelyn Smith, M.A.

Keywords: NDA review, clinical studies, noninferiority

Table of Contents

Table of Figures

Table of Tables

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Based on protocol specified and justified 10% non-inferiority margin, study A2309 demonstrated that both everolimus treatment regimens were non-inferior to the Myfortic treatment regimen at 12 months in the incidence rate of efficacy failure (composite of treated biopsy-proven acute rejection, graft loss, death or loss to follow-up). Additionally, the 12-month incidence of graft loss, death and loss to follow-up was similar between both everolimus groups and the Myfortic group, although numerically these events were more frequent in the everolimus groups compared to the Myfortic group.

As compared to the Myfortic treatment regimen, both everolimus treatment regimens were demonstrated to have similar renal function measured as estimated mean glomerular filtration rate (GFR) at 12 months post-transplantation.

There was a disproportionate rate of premature treatment discontinuation within 12 months in study A2309, driven by higher rates of adverse events in both everolimus groups compared to Myfortic. More patients in both of the everolimus groups prematurely discontinued study treatment and were subsequently switched to alternate therapy than in the Myfortic group, which may bias the interpretation of the study safety and efficacy results.

Analysis by gender revealed that among female patients, rates of premature treatment discontinuation, primary efficacy failure and graft loss and death were considerably higher in both everolimus groups compared to the Myfortic group.

1.2 Brief Overview of Clinical Studies

The original New Drug Application (NDA), 21-560, dated 12/23/2002, consisted of two Phase 3 controlled clinical trials (study B201 and study B251) in support of the safety and efficacy of everolimus (RAD) for prophylaxis of organ rejection in allogeneic kidney transplantation indication. The Agency took an approvable action (former regulatory terminology for a Complete Response) on the application based on findings of unacceptable renal toxicity of everolimus when given with full dose cyclosporine. In this re-submission, two therapeutic drug monitoring (TDM) regimens of everolimus given with reduced dose cyclosporine were evaluated in a single Phase 3 study.

The applicant submitted one new Phase 3, randomized, active-controlled, clinical trial in this application. Study A2309 entitled, “A 24-month, multicenter, randomized, open-label noninferiority study of efficacy and safety comparing two regimens of concentration-controlled everolimus in two doses (1.5 and 3.0 mg/day starting doses) with reduced Neoral versus 1.44 g Myfortic with standard dose Neoral in *de novo* renal transplant

recipients", was conducted to compare two regimens of concentration-controlled everolimus with reduced dose Neoral (cyclosporine) versus 1.44 g Myfortic with standard dose Neoral in *de novo* renal transplant recipients. Study A2309 enrolled 833 male or female renal recipients, between the ages of 18 and 70, who were undergoing primary kidney transplantation. Patients were randomized, in a 1:1:1 fashion, to 1.5 mg/day (0.75 mg bid) starting dose everolimus (dose was then adjusted to reach blood trough level target of 3-8 ng/mL) plus reduced dose Neoral, 3.0 mg/day (1.5 mg bid) starting dose everolimus (dose was then adjusted to reach blood trough level target of 6-12 ng/mL) plus reduced dose Neoral, or 1.44 g (0.72 g bid) Myfortic with standard dose Neoral.

The primary efficacy analysis was performed when all randomized patients completed 12 months of the study. Although the study duration is 24-month, the primary endpoint is based on the 12-month data, which serves as the basis of this re-submission for NDA 21-560. A secondary analysis of study A2309 will be performed when all randomized patients complete the full 24-month of study follow-up to collect more long term safety information on the treatment regimens.

The primary efficacy objective of study A2309 was to demonstrate that one or both everolimus treatment regimens are non-inferior with respect to the primary efficacy endpoint to active control Myfortic, based on a 10% non-inferiority margin. The sponsor provided a detailed justification for the 10% non-inferiority margin for the primary efficacy endpoint using historical information and mixed effects modeling. This justification was reviewed separately under IND 52,003, SN 919/SDN981 and submitted to DARRTS on 10/30/2009. The full review was included in Appendix 1.

The primary efficacy analysis was based on the intent-to-treat (ITT) population defined as all randomized patients. The primary efficacy analysis consisted of comparing the incidence of treated biopsy-proven acute rejection (BPAR) episode, graft loss, death, or loss to follow-up at 12 months post-transplant between the two everolimus regimens and the Myfortic regimen. To account for multiple comparisons, the Hochberg's procedure was used to maintain the overall type I error rate at $\alpha = 0.05$ level.

The primary safety objective of study A2309 was to demonstrate that comparable renal function, measured as GFR, is achieved between everolimus treatment regimens and Myfortic treatment regimen at 12-month post transplantation.

1.3 Statistical Issues and Findings

The primary efficacy endpoint of study A2309 was efficacy failure, defined as the composite consisting of treated BPAR episode, graft loss, death, or loss to follow-up at 12 months post-transplant. Using the protocol-defined Hochberg's procedure for multiple comparison adjustment, the study demonstrated that both of the everolimus treatment regimens were non-inferior to the Myfortic treatment regimen in the incidence of efficacy failure. The incidence rate of efficacy failure was 25.3%, 21.9% and 24.2% in the

everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. The difference between everolimus 1.5 mg and Myfortic was 1.1%, with 95% confidence interval (-6.1, 8.3). The difference between everolimus 3.0 mg and Myfortic was -2.3%, with 95% confidence interval (-9.3, 4.7). Both everolimus regimens were also demonstrated to be similar to the Myfortic regimen in the incidence of graft loss, death or loss to follow-up at 12 months (main secondary efficacy endpoint). Similar results were shown for other secondary efficacy endpoints.

Comparable values of estimated GFR at month 12 were shown between each of the everolimus regimens and the Myfortic regimen.

Premature treatment discontinuation, primarily due to adverse events, was frequent and statistically significantly higher in both everolimus groups, as compared to the Myfortic group. The incidence rate of premature treatment discontinuation was 30.0%, 34.7% and 21.7% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. The rates were statistically significantly higher in both everolimus groups (p=0.03, everolimus 1.5 mg, p=0.001 everolimus 3.0 mg) compared to the Myfortic group. The imbalanced incidence of treatment discontinuation should be considered, when interpreting the safety and efficacy outcomes of Study A2309 (more details could be found in Table 2, Figure 2, Section 3.1.2, 3.1.3 and 3.2.2). Confidence intervals obtained from a sensitivity analysis including premature treatment discontinuations as failures in the primary efficacy endpoint could not rule out that everolimus was no more than 10% worse than Myfortic.

Both the patient and the investigator were unblinded to the treatment regimen a patient received, because of the open-label design of study A2309. This must be taken into consideration in the interpretation of the study results, since unblinded studies are more subject to bias. This is of particular concern given the observed higher rates of premature treatment discontinuation in both everolimus groups, which may be related to the unblinded nature of the study.

The incidence of edema-related events was statistically significantly higher in both everolimus groups than in the Myfortic group (55.5% and 54.7% versus 45.1%, p-value was 0.02 and 0.03 respectively). Additionally, the incidence of any wound healing related events was 6.2%, 14.4% and 5.1% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively within 12 months post transplantation (p<0.001 everolimus 3.0 mg versus Myfortic, p=0.71 everolimus 1.5 mg versus Myfortic).

Subgroup analyses showed that efficacy results of study A2309 were not consistent across gender. There was a significant interaction noted between the everolimus 3.0 mg group and Myfortic by gender. The efficacy failure rate was lower in both everolimus groups than in the Myfortic group in male patients. Among female patients, the efficacy failure rate and the rate of graft loss and death were higher in both everolimus groups than Myfortic. Furthermore, for female patients, incidence of premature treatment discontinuation was significantly higher in each of the everolimus groups than in the Myfortic group. Study A2309 may not provide adequate information to determine a safe

and efficacious everolimus regimen for females (more details could be found in Table 18, Table 19 and Table 21, Section 4.1). No differences were seen in the subgroup analyses by age (< 50 years and ≥ 50 years) or by race (Black versus non-Black), although only about 14% of patients were Black, the others were Caucasian, Asian and other races.

2. INTRODUCTION

2.1 Overview

Everolimus, a member of the mTOR inhibitor class of immunosuppressants, is a hydroxyethyl derivative of rapamycin, was submitted as a new immunosuppressive agent. The proposed indication is prophylaxis of acute rejection in *de novo* renal transplantation.

The original NDA submitted on 12/23/2002 consisted of two Phase 3, randomized, controlled clinical trials (study B201 and study B251) in support of the safety and efficacy of everolimus for prophylaxis of organ rejection in allogeneic kidney transplantation indication. The Agency issued an approvable action (former regulatory action for a Complete Response) stating concerns regarding unacceptable renal toxicity of fixed dose everolimus when given with full dose cyclosporine. The applicant subsequently designed and conducted a new Phase 3, randomized clinical trial, A2309, to assess two therapeutic drug monitoring (TDM) regimens of everolimus when given with reduced dose cyclosporine in *de novo* kidney transplant patients. This document will review the efficacy and safety of this new study. Please see the statistical review by Dr. Ruthanna Davi dated 10/16/2003 for a review of the previous two studies (study B201 and B251). Additionally, please refer to the statistical review by Dr. John Yap for a more detailed review of certain safety aspects of the study A2309.

Study A2309 entitled, "A 24-month, multicenter, randomized, open-label noninferiority study of efficacy and safety comparing two regimens of concentration-controlled everolimus in two doses (1.5 and 3.0 mg/day starting doses) with reduced Neoral versus 1.44 g Myfortic with standard dose Neoral in *de novo* renal transplant recipients" was a prospective, 24-month, multicenter, randomized, open-label, noninferiority study in *de novo* renal transplant recipients. A total of 833 patients were enrolled in approximately 80 clinical sites located in 16 countries and regions. The primary 12-month data analysis of Study A2309 was submitted in this re-submission. A secondary analysis, to access long-term safety, will be conducted when all randomized patients complete 24 months follow-up. For the 12-month data analysis, the cut-off day was the date of Month 12 visit or Day 450 whichever occurred earliest.

2.2 Data Sources

The applicant submitted electronic documents and datasets for study A2309. The following files available within the CDER Electronic Document Room (EDR) were utilized in this review.
Data sources include all material reviewed, e.g. applicant study reports, data sets analyzed, and literature referenced.

\\Cdsesub1\evsprod\NDA021560\0010\m5\datasets\rad001a2309
\\Cdsesub1\evsprod\NDA021560\0016\m5\datasets\rad001a2309

Reviewer requested data files
\\Cdsesub1\evsprod\NDA021560\0017\m5\datasets\rad001a2309\analysis
\\Cdsesub1\evsprod\NDA021560\0019\m5\53-clin-stud-rep\535-rep-effic-safety-stud\prophylaxis-of-kidney-transplant-rejection\5351-stud-rep-contr\rad001a2309
\\Cdsesub1\evsprod\NDA021560\0020\m5\53-clin-stud-rep\535-rep-effic-safety-stud\prophylaxis-of-kidney-transplant-rejection\5351-stud-rep-contr\rad001a2309
\\Cdsesub1\evsprod\NDA021560\0021\m5\datasets\rad001a2309\analysis
\\Cdsesub1\evsprod\NDA021560\0023\m5\datasets\rad001a2309\analysis
\\Cdsesub1\evsprod\NDA021560\0026\m5\datasets\rad001a2309\analysis

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design and Endpoints

Study A2309 was a 24-month, multicenter, randomized, open-label, active controlled study, as shown in Figure 1. Patients were randomized (stratified by center), within 24 hours after kidney transplantation surgery, in a 1:1:1 fashion, to one of the following treatment groups:

Group 1: 1.5 mg (0.75 mg bid) everolimus + Simulect + Neoral +/- corticosteroids

Group 2: 3.0 mg (1.5 mg bid) everolimus + Simulect + Neoral +/- corticosteroids

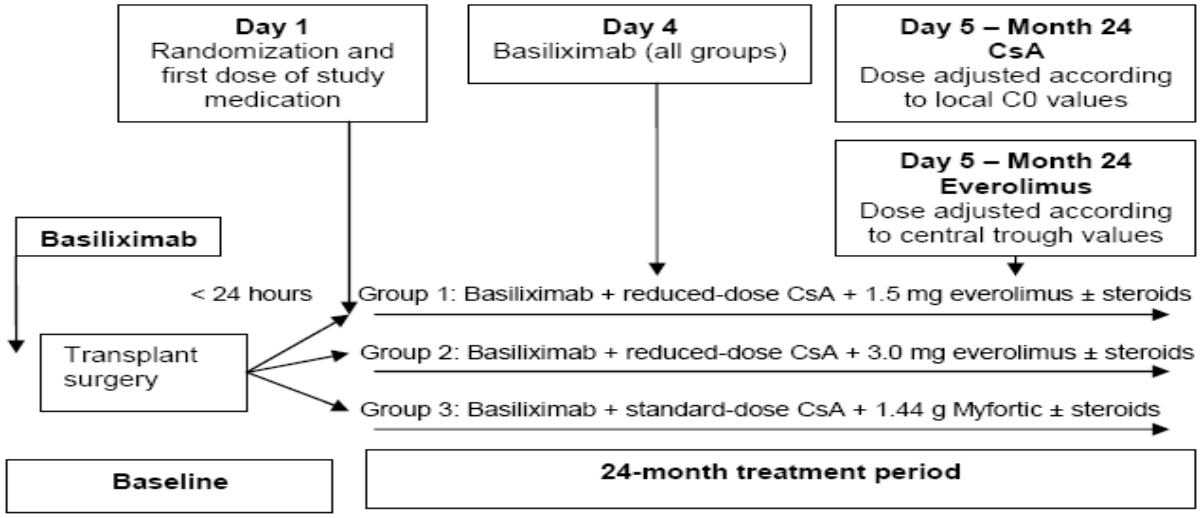
Group 3: 1.44 g (0.72 bid) Myfortic + Simulect + Neoral +/- corticosteroids

Therapeutic drug monitoring was required throughout the duration of the study in the everolimus groups to achieve target everolimus trough concentrations. On day 5, the 0.75 mg bid dose will be increased if the trough level was <3 ng/mL, and reduced if >8 ng/mL on 2 consecutive measures. The 1.5 mg bid dose was increased if the trough level was < 6 ng/mL, and reduced if >12 ng/mL on 2 consecutive measures.

In all three treatment regimens, Neoral dosing was managed using cyclosporine trough level (C0). In both everolimus groups, Neoral dose was adjusted to get C0 value within

the pre-specified target ranges: Starting at the Day 1 visit: 100-200 ng/mL, starting at the Month 2 visit : 75-150 ng/mL, starting at the Month 4 visit: 50-100 ng/mL and starting at the Month 6 visit: 25-50 ng/mL. In the Myfortic group, Neoral dose was adjusted to get C0 value within the following range for the time of the study: Starting at the Day 1 visit: 200-300 ng/mL, starting at the Month 2 visit and thereafter: 100-250 ng/mL.

Figure 1: Study Design of A2309



Note: Figure obtained from applicant's clinical study report, page 96

If an investigator chose to stop cyclosporine or to add another immunosuppressive agent to the regimen, study medication was discontinued for the patient. However, the patient continued to be followed until completion of the trial.

Inclusion criteria consisted of male and female patients, 18-70 years of age, who underwent primary cadaveric, living unrelated or non HLA identical living related kidney transplantation.

The primary efficacy endpoint was efficacy failure at 12 months post transplantation, where efficacy failure is the composite endpoint consisting of treated biopsy-proven acute rejection (BPAR) episode (based on local laboratory assessment), graft loss, death, or loss to follow-up.

The main secondary efficacy endpoint was the incidence of composite of graft loss, death or loss to follow-up at 12 months. Other secondary endpoints included primary efficacy endpoint at 6 months, treated BPAR at 6 and 12 months, graft loss at 6 and 12 months, death at 6 and 12 months, biopsy proven chronic allograft nephropathy (CAN) at 12 months, graft loss or death at 6 and 12 months, graft loss, death or loss to follow up at 6 months, and antibody treated BPAR at 12 months. Additionally, sensitivity analysis of

the primary efficacy endpoint was conducted in which the treated BPAR endpoint was based on central laboratory assessment.

A main secondary safety endpoint was calculated GFR [using the modification of diet in renal disease (MDRD) formula] at 12 months. The safety objective was to demonstrate that the mean GFR for the everolimus groups is not lower (not worse) than that of the Myfortic group by 8 ml/min/1.73m^2 or more.

Note: It is atypical to specify a NI margin for a safety variable particularly given the difficulty in justifying the NI margin. Therefore, the Division will consider results from the safety analysis along with efficacy in an overall risk benefit assessment.

After baseline, scheduled visit for patients are at Day 1, 3, 4, 5, 7, 14, 28, and afterwards at Week 2, 3, 4, 6, 7, 9, 12, 18, 24. No deviation in the evaluation schedule should happen during Days 1 through 7. After day 7, a visit window of 2 days up to Day 28, 1 week between Day 28 and Month 6, and 2 weeks after Month 6 is acceptable.

All patients discontinuing the study medication prior to the 24-month treatment period were contacted at scheduled months 3, 6, 9, 12, 18 and 24 visits to obtain follow-up information and should not be considered withdrawn from the study. Information was collected on rejection episodes, proteinuria, serum creatinine, graft loss/re-transplant, SAEs, malignancies, opportunistic infections, patient survival and immunosuppressive therapy. In addition, major adverse cardiac events (MACE) would be reported during the follow-up period.

Note: Adverse events and serious adverse events were systematically collected up to 7 and 30 days, respectively, following premature treatment discontinuation.

3.1.2 Statistical Methods

All efficacy analyses were performed using the ITT population, defined as all randomized patients. Safety analyses were performed using the safety population defined as all randomized patients who received at least one dose of study medication and who had at least one safety assessment.

The primary efficacy objective of study A2309 was to demonstrate that one or both everolimus regimens were non-inferior to the Myfortic regimen, based on a pre-specified 10% non-inferiority margin, in incidence of the primary efficacy failure endpoint. The applicant previously submitted a justification for the 10% non-inferiority margin for the primary efficacy endpoint of efficacy failure at 12 months (IND 52,003 SN919/SDN981, dated 6/29/2009). A summary of this justification is given in Appendix 1.

The primary null hypothesis is as follows:

H_0= $\theta_c - \theta_m \geq 0.10$ (non-inferiority margin): the proportion (θ_c) of patients experiencing efficacy failure at 12 months on the everolimus group is higher than that of the Myfortic group (θ_m) by 10% or more.

The alternative hypothesis is as follows:

H_1= $\theta_c - \theta_m < 0.10$: the proportion ($\theta_c$) of patients experiencing efficacy failure at 12 months on the everolimus group is not higher than that of the Myfortic group (θ_m) by 10% or more.

The null hypothesis was tested by means of the z-statistic using a two-sided 0.025 significance level.

In order to adjust for multiple comparisons, Hochberg's procedure was used to maintain the overall type I error rate at $\alpha = 0.05$ level. If the upper bound of the 95% confidence intervals around the differences in proportions of efficacy failure between each everolimus group and Myfortic group were within the 10% non-inferiority margin, both regimens of everolimus were considered non-inferior to Myfortic 1.44 g regimen. Otherwise, each pair-wise comparison (i.e. everolimus v. Myfortic) was considered separately using a 97.5% confidence interval. Time to efficacy failure was calculated by using Kaplan-Meier (KM) methodology and compared by using the log-rank test. The Breslow-Day test, with a 0.10 significance level, was used to test the interaction between treatment groups and study centers.

Main secondary endpoint (the composite of graft loss, death or loss to follow-up) was compared using the same non-inferiority margin as the primary efficacy endpoint.

Note: Because of the lack of information, it is difficult to justify the non-inferiority margin for the secondary endpoints. Please refer to the statistical review of IND 52,003 for more detailed discussion.

The main safety objective of study A2309 was to demonstrate that either one or both of the everolimus groups were comparable to the Myfortic group with respect to mean GFR at 12 months.

The null hypothesis is as follows:

The mean GFR of the everolimus group is lower (worse) than that of the Myfortic group by 8 mL/min/1.73m^2 or more.

The alternative hypothesis is as follows:

The mean GFR for the everolimus group is not lower (not worse) than that of the Myfortic group by 8 mL/min/1.73m^2 or more.

Note: It is atypical to specify a NI margin for a safety variable particularly given the difficulty in justifying the NI margin. Therefore, the Division will consider results from the safety analysis along with efficacy in an overall risk benefit assessment.

The following imputation methods were applied for patients with missing 12-month GFR value:

1. Patients who lost their graft were assigned a value zero for their 12-month GFR value.
2. Patients who died with a functioning graft had an imputed 12-month GFR value using the last-on-treatment-observation-carried- forward (LOCF) method.
3. Patients who had no GFR value past Day 315 (Day 316 being the lower limit of the 12-month visit window) and had neither died nor had graft loss would have an imputed 12-month GFR value using the LOCF method.

Additionally, analyses were conducted using different imputation methods. Refer to the statistical safety review by Dr. John Yap for details on these analyses.

For study A2309, there was no planned formal interim analysis (IA) prior to the 12-month database lock. A data monitoring committee was reviewing data every 6 months for safety monitoring. Six-month data will be analyzed using the locked 12-month database.

As a sensitivity analysis to the primary analysis, the analysis was repeated for the composite efficacy endpoint in which the treated BPAR endpoint was derived from the central pathologist(s) readings. The discordance between local and central biopsy readings was assessed by treatment group for all paired readings using Kappa Coefficient (Kappa Statistic) along with asymptotic standard error and 95% CI.

The sample size calculation is based on the primary endpoint with the assumption that the efficacy failure rate is 20% for the Myfortic group and 19% for the everolimus groups. Based on these assumptions a sample size of 275 patients per group will have 84% power to show that the everolimus group is not more than 10% worse than the Myfortic group with respect to the 12-month composite efficacy failure rate.

3.1.3 Patient Disposition, Demographic and Baseline Characteristics

As shown in Table 1, baseline demographics and characteristics for recipients and donors were similar among the three treatment groups. All randomized patients were between the ages of 18 and 70 years; more than 43% were 50 years of age or older. More than 63% of all recipients were male and more than 64% were Caucasian. Among donors, more than 49% were male and more than 68% were Caucasian. The primary diseases leading to end-stage renal disease in recipients was similar across the treatment groups. The most frequent diseases leading to transplantation were hypertension/nephrosclerosis (18.1%), glomerulonephritis/glomerular disease (16.6%) and diabetes mellitus (13.6%). There were no major differences among treatment groups with regards to recipient disease characteristics.

Table 1: Baseline Demographics by Treatment Group (ITT Population)

	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Recipient Age (years)			
Mean (± SD)	45.7 (12.7)	45.3 (13.4)	47.2 (12.7)
Range	18 – 70	18 – 70	18 – 70
Donor Age (years)			
Mean (± SD)	41.4 (13.9)	41.1 (13.0)	41.8 (13.6)
Range	5 – 67	5 – 69	5 – 67
Recipient age group, n (%)			
< 50 years	156 (56.3)	153 (54.8)	143 (51.6)
≥ 50 years	120 (43.3)	126 (45.2)	134 (48.4)
Unknown	1 (0.4)	0 (0)	0 (0)
Donor age group, n (%)			
< 50 years	181 (65.3)	203 (72.8)	182 (65.7)
≥ 50 years	95 (34.3)	76 (27.2)	94 (33.9)
Unknown	1 (0.4)	0 (0)	1 (0.4)
Recipient gender, n (%)			
Male	176 (63.5)	191 (68.5)	189 (68.2)
Female	100 (36.1)	88 (31.5)	88 (31.8)
Unknown	1 (0.4)	0 (0)	0 (0)
Donor gender, n (%)			
Male	154 (55.6)	139 (49.8)	136 (49.1)
Female	122 (44.0)	140 (50.2)	140 (50.5)
Unknown	1 (0.4)	0 (0)	1 (0.4)
Recipient race, n (%)			
Caucasian	193 (69.7)	180 (64.5)	190 (68.6)
Black	34 (12.3)	40 (14.3)	38 (14.1)
Asian	32 (11.6)	38 (13.6)	36 (13.0)
Other	17 (6.1)	21 (7.5)	12 (4.3)
Unknown	1 (0.4)	0 (0)	0 (0)
Donor race, n (%)			
Caucasian	193 (69.7)	191 (68.5)	197 (71.1)
Black	20 (7.2)	22 (7.9)	25 (9.0)
Asian	32 (11.6)	35 (12.5)	31 (11.2)
Other	27 (9.8)	26 (9.3)	19 (6.9)
Unknown	5 (1.8)	5 (1.8)	5 (1.8)
Recipient BMI (kg/m^2)			
Mean (± SD)	25.8 (5.1)	25.8 (5.0)	25.9 (4.7)
Range	15.7 – 43.6	15.2 – 39.5	17.2 – 42.3
Primary Disease for Transplantation, n (%)			
Hypertension/nephrosclerosis	50 (18.1)	56 (20.1)	45 (16.3)
Glomerulonephritis/glomerular disease	43 (15.5)	55 (19.7)	40 (14.4)

Diabetes mellitus	39 (14.1)	29 (10.4)	45 (16.3)
Polycystic disease	36 (13.0)	29 (10.4)	33 (11.9)
IgA nephropathy	18 (6.5)	17 (6.1)	29 (10.5)
Other	56 (20.2)	56 (20.1)	45 (16.2)
Unknown	35 (12.6)	37 (13.3)	40 (14.4)

Among the 833 randomized patients in the ITT population, approximately 29% prematurely discontinued study medication by Day 450, which was the protocol defined cutoff date for 12 month analyses. As presented in Table 2, there was an imbalance across treatment groups in the incidence of premature treatment discontinuation. At Month 12, the incidence of premature treatment discontinuation in the everolimus 1.5 mg, 3.0 mg and Myfortic groups was 30.0% (83/277), 34.1% (95/279), and 21.7% (60/277) respectively. Compared to the Myfortic group, the incidence was statistically significantly higher in the everolimus 1.5 mg group (p-value=0.03, Fisher's exact test) and in the everolimus 3.0 mg group (p-value=0.001, Fisher's exact test).

The most common reason reported for premature discontinuation of study treatment was adverse events, which accounted for 18%, 20%, and 9% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. In the everolimus 1.5 mg group, 18.1% of the patients prematurely discontinued treatment due to adverse events, which was statistically significantly higher than the incidence in the Myfortic group (9.4%) with p-value=0.004 (Fisher's exact test). The incidence of treatment discontinuation due to adverse events was also statistically significantly higher in the everolimus 3.0 mg group than in the Myfortic group (20.4% versus 9.4%, with p-value<0.0001, Fisher's exact test).

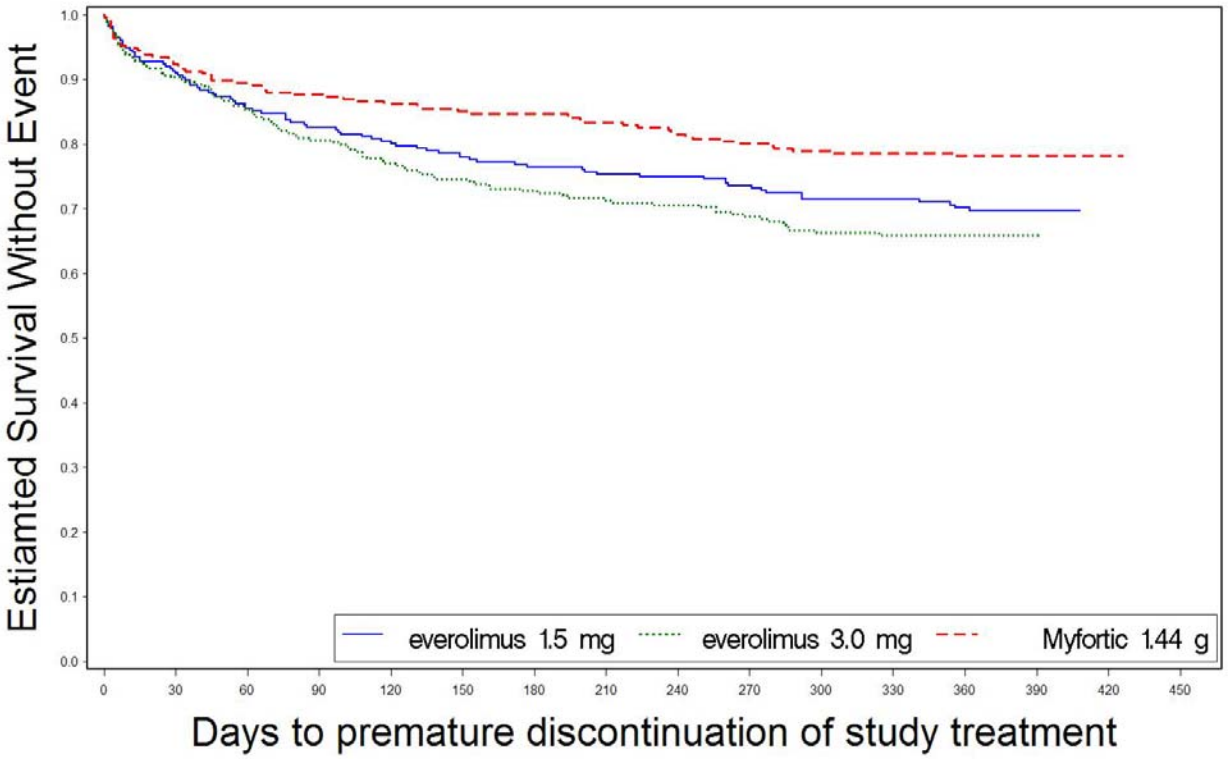
Approximately 12% of the randomized patients prematurely discontinued the study phase. Study discontinuations were more frequent in both of the everolimus groups compared to the Myfortic group (13.7% and 11.8% versus 10.1%), but the differences were not statistically significant (p-value= 0.24 and 0.59 respectively, Fisher's exact test). There did not appear to be one primary reason for premature study discontinuation occurring more frequently in the everolimus groups compared to the Myfortic group.

Table 2: Premature Study Medication or Study Phase Discontinuation by Treatment Group (ITT Population - 12 Month Analysis)

Number of patients (%)	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Discontinued study medication	**83 (30.0)**	**95 (34.1)**	**60 (21.7)**
Adverse event(s)	50 (18.1)	57 (20.4)	26 (9.4)
Unsatisfactory therapeutic effect	11 (4.0)	14 (5.0)	13 (4.7)
Subject withdrew consent	11 (4.0)	4 (1.4)	5 (1.8)
Graft loss	3 (1.1)	6 (2.2)	6 (2.2)
Death	3 (1.1)	3 (1.1)	4 (1.4)
Protocol deviation	2 (0.7)	5 (1.8)	2 (0.7)
Abnormal lab value	1 (0.4)	4 (1.4)	1 (0.4)
Administrative problems	2 (0.7)	1 (0.4)	2 (0.7)
Abnormal test procedure	0 (0)	1 (0.4)	0 (0)
Unknown	0 (0)	0 (0)	1 (0.4)
Discontinued study phase	**38 (13.7)**	**33 (11.8)**	**28 (10.1)**
Subject withdrew consent	20 (7.2)	8 (2.9)	12 (4.3)
Graft loss	9 (3.3)	10 (3.6)	7 (2.5)
Death	7 (2.5)	9 (3.2)	6 (2.2)
Unknown	2 (0.7)	6 (2.2)	3 (1.1)

The Kaplan Meier plot for the time-to-event analysis of premature treatment discontinuation is shown in Figure 2. In both everolimus groups, treatment discontinuation occurred earlier and more often than the treatment discontinuation in the Myfortic group (p-value of log-rank test was 0.03 for everolimus 1.5 mg versus Myfortic, and was 0.002 for everolimus 3.0 mg versus Myfortic). These trends remained throughout the 12-month follow-up period, as the KM curves were proportional.

Figure 2: Kaplan-Meier for Premature Treatment Discontinuation by Treatment Group (ITT Population - 12 Month Analysis)



3.1.4 Primary Efficacy Results

The primary efficacy endpoint of study A2309 was a composite consisting of the first occurrence of treated BPAR (local assessment), death, graft loss, or loss to follow-up measured at 12 months following kidney transplantation. Based on the ITT population, 25.3%, 21.9%, and 24.2% of the patients experienced efficacy failure within 12 months, in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively (Table 3). The difference in failure rates between everolimus 1.5 mg and Myfortic was 1.1% with 95% CI (-6.1, 8.3) and between 3.0 mg and Myfortic was -2.3% with 95% CI (-9.3, 4.7).

Based on the protocol defined and justified non-inferiority margin of 10% and using the Hochberg's procedure to adjust for multiple comparisons, non-inferiority of both everolimus groups to Myfortic with respect to the primary efficacy endpoint was achieved. This was demonstrated by the fact that the upper limits of both 95% confidence intervals were less than the 10% non-inferiority margin.

Table 3: Primary Efficacy Endpoint Analysis by Treatment Group (ITT Population - 12 Month Analysis)

Number of patients (%)	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Efficacy Failure	**70 (25.3)**	**61 (21.9)**	**67 (24.2)**
Treated BPAR	45 (16.3)	37 (13.3)	47 (17.0)
Graft Loss	12 (4.3)	13 (4.7)	9 (3.3)
Death	7 (2.5)	10 (3.6) *	6 (2.2)
Loss to follow-up	12 (4.3)	8 (2.9) **	9 (3.3)
95% CI (everolimus-Myfortic)	**(-6.1, 8.3)**	**(-9.3, 4.7)**	N/A
97.5% CI (everolimus-Myfortic)	(-7.1, 9.3)	(-10.3, 5.7)	N/A

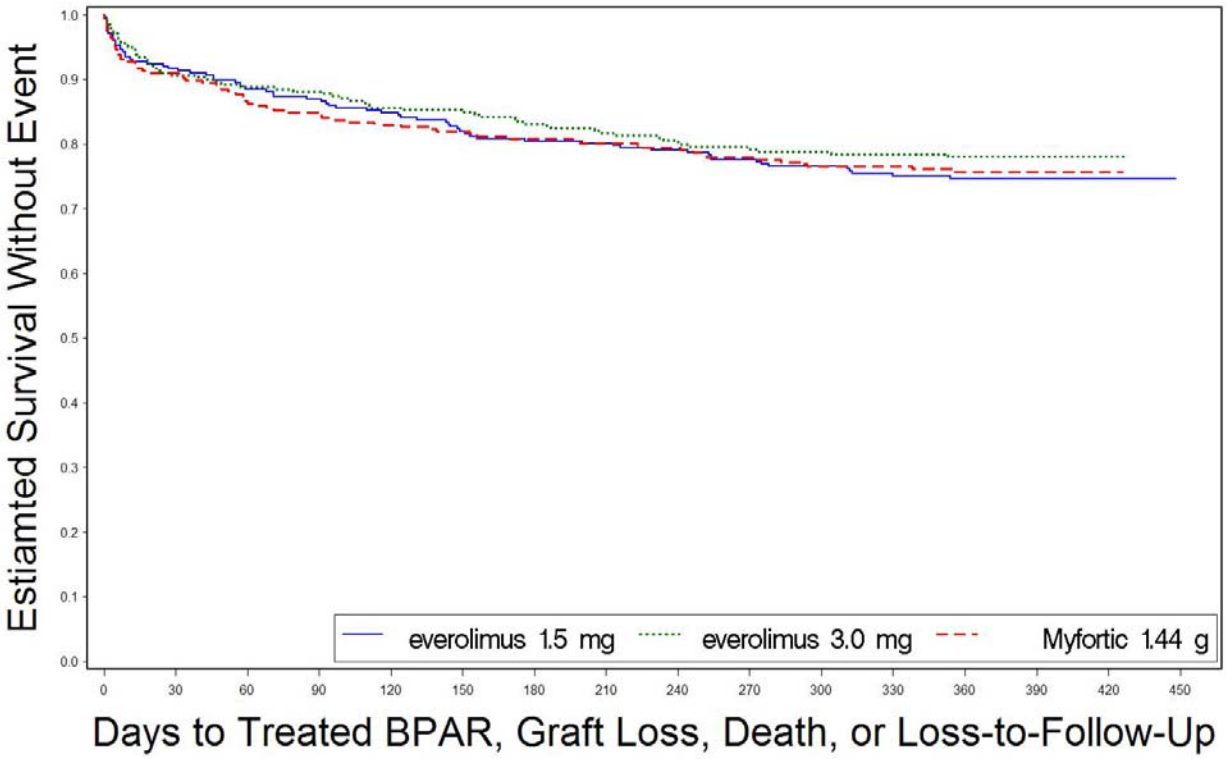
* One patient who died 10 days after withdrew consent was included

** One patient who had graft loss before the randomization was considered as loss to follow-up

The Kaplan Meier plot for the primary efficacy endpoint within 12 months is provided in Figure 3. Based on the log-rank test, median time to event was not statistically significantly different between everolimus 1.5 mg (p-value=0.83) and everolimus 3.0 mg (p-value=0.49) and Myfortic. Additionally, no statistically significant differences in time to event were shown between the two everolimus groups (p-value=0.37). If loss to follow-up patients were treated as censored rather than efficacy failure, similar results were reported by using time-to-event analyses. The p-value of the log-rank test was 0.97 for everolimus 1.5 mg group versus Myfortic 1.44 g group, and was 0.53 for everolimus 3.0 mg group versus 1.44 g group, demonstrating that no significant differences between each of the everolimus groups and the Myfortic group were seen.

Note: The division always considers loss to follow-up as failure in renal transplantation.

Figure 3: Kaplan-Meier Estimates for the Primary Efficacy Endpoint by Treatment group (ITT Population - 12 Month Analysis)



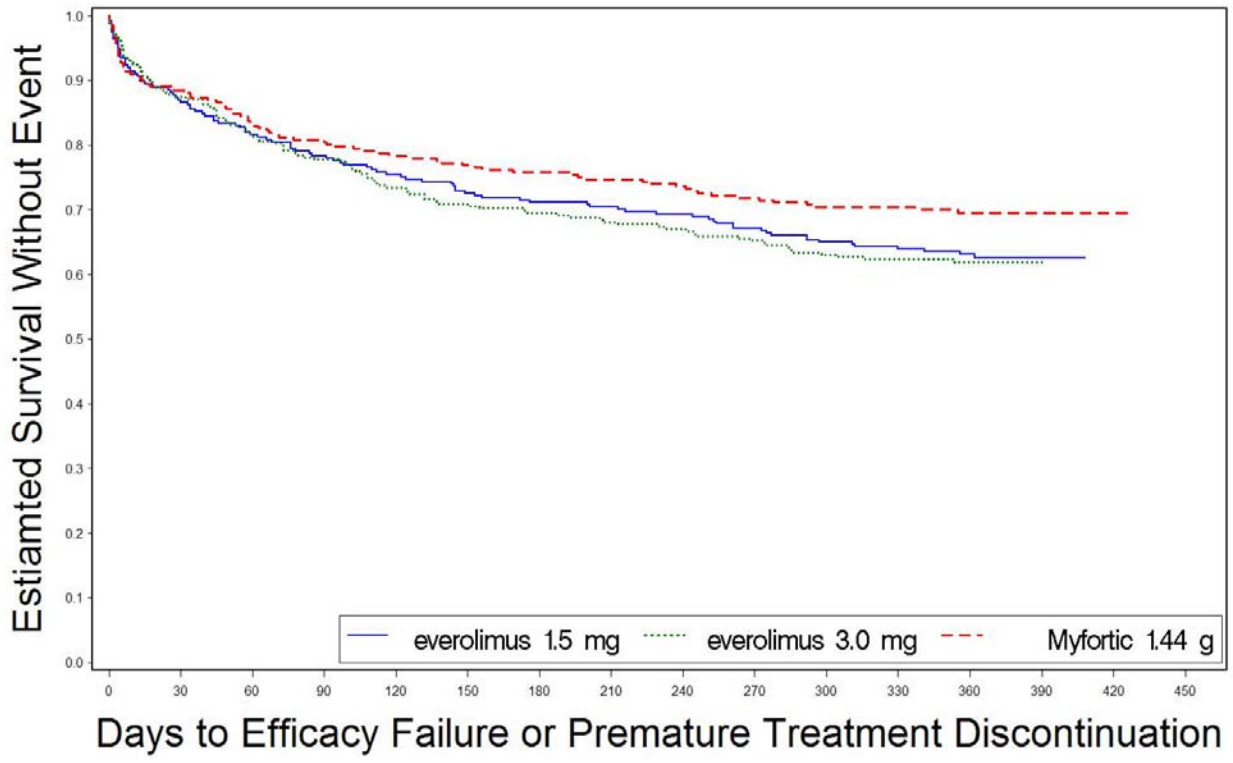
To assess the impact of the disproportionate rates of premature treatment discontinuation on the primary efficacy endpoint, treatment discontinuation was treated as failure along with the primary efficacy composite endpoint. In this sensitivity analysis, the upper bounds of the 95% and 97.5% confidence intervals for the differences of both everolimus groups compared to Myfortic exceeded 10% (Table 4). The justification for the 10% non-inferiority margin did not account for premature treatment discontinuation; however, if rates of premature treatment discontinuation were similar among treatment regimens in historical clinical trials, a 10% margin including this event in not unreasonable. Kaplan Meier plot for this sensitivity analysis is provided in Figure 4. The survival curve of the Myfortic group was consistently higher than both everolimus groups; however, the median times to event were not statistically significantly different based on log-rank test (p= 0.12 for everolimus 1.5 mg versus Myfortic, and p=0.08 for everolimus 3.0 mg versus Myfortic).

Table 4: Primary Efficacy Endpoint with Premature Treatment Discontinuation as Failure by Treatment Group (ITT Population - 12 Month Analysis)

Number of patients (%)	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Efficacy Failure or Premature Treatment Discontinuation *	103 (37.2)	106 (38.0)	84 (30.3)
95% CI (everolimus - Myfortic)	(-1.0, 14.7)	(-0.2, 15.5)	N/A
97.5% CI (everolimus - Myfortic)	(-2.1, 15.8)	(-1.3, 16.7)	N/A

* Sensitivity analysis were failure is defined as treated BPAR, death, graft loss, loss to follow-up or premature treatment discontinuation by month 12 post-transplantation

Figure 4: Kaplan-Meier Estimates for Primary Efficacy Failure or Premature Treatment Discontinuation by Treatment Group (ITT Population - 12 Month Analysis)



3.1.5 Secondary Efficacy Results

The main secondary efficacy objective of study A2309 was to compare the incidence rate of the composite of graft loss, death, or loss to follow-up between the everolimus and

Myfortic groups at 12 months post-transplantation. As presented in Table 5, the incidence of graft loss, death or loss to follow-up was similar between the two everolimus groups, (11.6% and 11.1% respectively), and was slightly lower in the Myfortic group (9.4%). The difference in failure rates between everolimus 1.5 mg and Myfortic was 2.2% with 95% CI (-2.9, 7.3) and between 3.0 mg and Myfortic was 1.7% with 95% CI (-3.3, 6.8). The applicant stated that a 10% margin would be used and both 95% confidence intervals for the everolimus groups compared to Myfortic excluded this margin based on the upper bound.

Note that there is not a justified non-inferiority margin for the endpoint of death, graft loss or loss to follow-up. This is due to the lack of historical information to derive an estimate of the treatment effect of active control over placebo for death, graft loss and loss to follow-up events.

As presented in Figure 5, the Kaplan Meier plots for the failure event of graft loss, death, or loss to follow-up within 12 months were similar between each of the everolimus groups and the Myfortic group. The p-value of log-rank test was 0.40 for everolimus 1.5 mg group versus Myfortic 1.44 g group, and was 0.51 for everolimus 3.0 mg group versus Myfortic 1.44 g group.

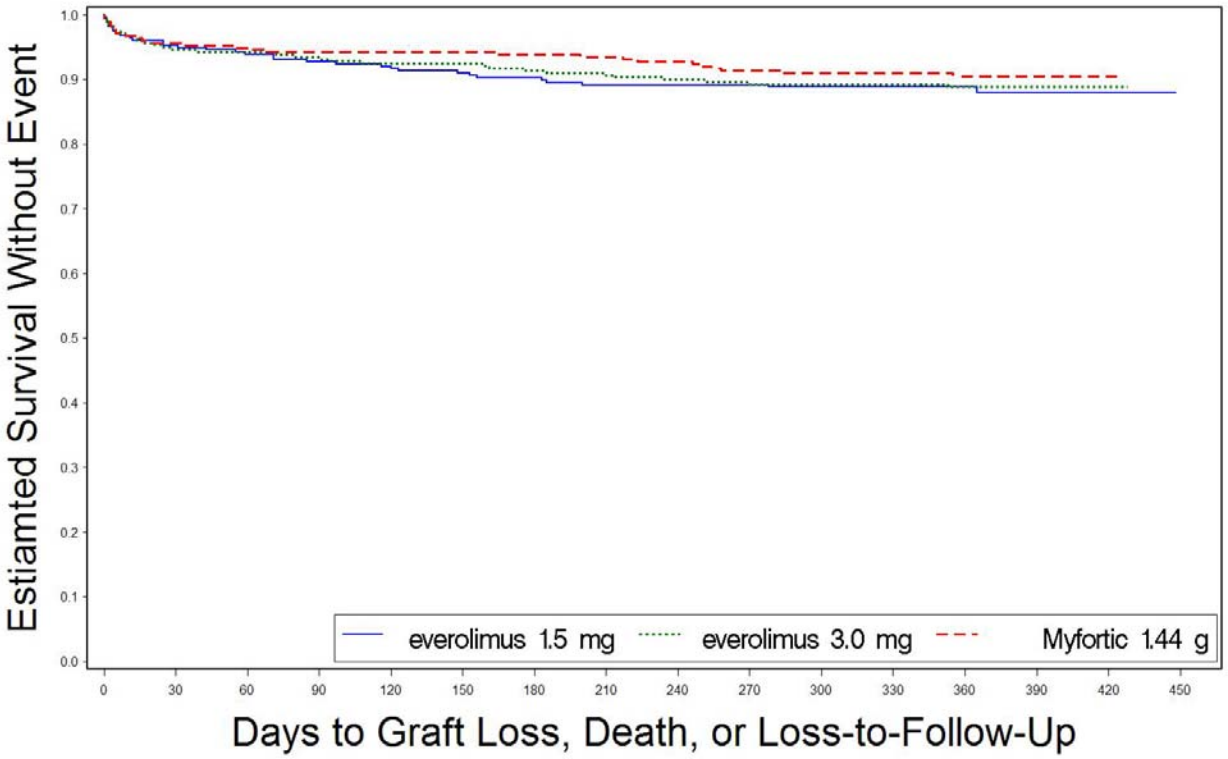
Table 5: Main Secondary Efficacy Endpoint Analysis by Treatment Group (ITT Population - 12 Month Analysis)

Number of patients (%)	**everolimus 1.5 mg (N=277)**	**everolimus 3.0 mg (N=279)**	**Myfortic 1.44 g (N=277)**
Graft loss, death or loss to follow up	**32 (11.6)**	**31 (11.1)**	**26 (9.4)**
Graft Loss	12 (4.3)	13 (4.7)	9 (3.3)
Death	7 (2.5)	10 (3.6) *	6 (2.2)
Loss to follow-up **	14 (5.1)	10 (3.6)	11 (4.0)
95% CI (everolimus-Myfortic)	**(-2.9, 7.3)**	**(-3.3, 6.8)**	N/A
97.5% CI (everolimus-Myfortic)	(-3.7, 8.0)	(-4.0, 7.5)	N/A

* One patient who died 10 days after withdrew consent was included

** A loss to follow-up patient was a patient who did not experience graft loss or death and whose last day of contact is prior to study Day 316

Figure 5: Time to Graft Loss, Death, or Loss to Follow-up by Treatment Group (ITT Population - 12 Month Analysis)



Analyses of Treated BPAR Events

As presented in Table 6 and Table 7, the grades and the number of treated BPAR, based on local assessment, were similar between each of the everolimus groups and the Myfortic group. In all three treatment groups, more than 85% of patients who experienced treated BPAR had only one treated BPAR event and the majority of BPAR were of Banff Type IA-the least severe grade. Few patients experienced more severe BPAR of Type III.

Table 6: Grade of Treated BPAR by Treatment Group (ITT Population - 12 Month Analysis)

	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Number of patient (%) with any grade of treated BPAR *	**45 (16.3)**	**37 (13.3)**	**47 (17.0)**
Banff Type IA	21 (7.6)	16 (5.7)	22 (7.9)
Banff Type IB	7 (2.5)	9 (3.2)	6 (2.2)
Banff Type IIA	7 (2.5)	9 (3.2)	15 (5.4)
Banff Type IIB	1 (0.4)	3 (1.1)	2 (0.7)
Banff Type III	1 (0.4)	0 (0)	1 (0.4)
Missing grade	6 (2.2)	4 (1.4)	3 (1.1)

* Based on local pathology assessment

Table 7: Number of Treated BPAR by Treatment Group (ITT Population - 12 Month Analysis)

Number of patient (%) with treated BPAR by number of BPAR *	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
0 treated BPAR	232 (83.8)	242 (86.7)	230 (83.0)
1 treated BPAR	39 (14.1)	32 (11.5)	41 (14.8)
2 treated BPAR	5 (1.8)	5 (1.8)	5 (1.8)
3 treated BPAR	0 (0)	0 (0)	0 (0)
4 treated BPAR	1 (0.4)	0 (0)	0 (0)

* Based on local pathology assessment

The discordance between local and central assessments of treated BPAR was analyzed using the Kappa coefficient (Table 8). The Kappa coefficients were comparable across treatment groups (0.44, 0.47, and 0.52 for everolimus 1.5mg, 3.0mg and Myfortic groups respectively), indicating that readings from the local and central laboratories were moderate concordant (full concordance is when Kappa=1).

Note: The number of patient with treated BPAR based on central assessment was noticeably smaller, compared to the number based on local assessment. This was due to the fact that a considerable amount of local biopsies did not undergo a central reading. The local assessment is viewed as primary for determination of treated BPAR as it is this assessment that investigators use in therapy management. Central biopsy assessments often occur later after the kidney biopsy is obtained and are therefore considered as secondary.

Table 8: Concordance between Local and Central Assessment of Treated BPAR (ITT Population - 12 Month Analysis)

Number of patient (%) with treated BPAR	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
Based on local assessment	45 (16.3)	37 (13.3)	47 (17.0)
Based on central assessment	20 (7.2)	14 (5.0)	20 (7.2)
Kappa coefficient (Standard Error)	0.44 (0.08)	0.47 (0.09)	0.52 (0.07)
95% CI for Kappa coefficient	(0.28, 0.59)	(0.30, 0.64)	(0.37, 0.66)

There were no significant differences between the everolimus groups and the Myfortic group in other secondary efficacy endpoints, including the incidence of efficacy failure within 6 months, graft loss or death at 6 and 12 months, graft loss, death or loss to follow-up at 6 months, and antibody treated BPAR at 12 months. These results are presented in Table 9.

Note: These analyses are not adjusted for multiple comparisons.

Table 9: Other Secondary Efficacy Endpoints Analysis by Treatment Group (ITT Population)

Number of patients (%)	everolimus 1.5 mg (N=277)	everolimus 3.0 mg (N=279)	Myfortic 1.44 g (N=277)
6-months Results			
Efficacy Failure	52 (18.8)	47 (16.9)	52 (18.8)
Treated BPAR	30 (10.8)	28 (10.0)	38 (13.7)
Graft Loss	11 (4.0)	11 (3.9)	8 (2.9)
Death	6 (2.2)	6 (2.2) *	3 (1.1)
Loss to follow-up **	9 (3.3)	7 (2.5)	5 (1.8)
95% CI (everolimus-Myfortic)	(-6.5, 6.5)	(-8.3, 4.4)	N/A
Graft loss or death	16 (5.8)	16 (5.7)	11 (4.0)
95% CI (everolimus-Myfortic)	(-1.8, 5.4)	(-1.8, 5.3)	N/A
Graft loss, death, or loss to follow-up	26 (9.4)	23 (8.2)	16 (5.8)
95% CI (everolimus-Myfortic)	(-0.8, 8.0)	(-1.8, 6.7)	N/A
12-months Results			
Graft loss or death	18 (6.5)	21 (7.5)	15 (5.4)
95% CI (everolimus-Myfortic)	(-2.9, 5.0)	(-2.0, 6.2)	N/A
Antibody treated BPAR	10 (3.6)	12 (4.3)	15 (5.4)
95% CI (everolimus-Myfortic)	(-5.3, 1.7)	(-4.7, 2.5)	N/A

* One patient who died 10 days after withdrew consent was included

** A loss to follow-up patient was a patient who did not experience treated BPAR, graft loss or death and whose last day of contact is prior to study Day 151. Therefore, one patient who was reported as loss to follow-up on Day 153 was not counted as loss to follow-up in 6 month analysis.

3.2 Evaluation of Safety

The main safety objective in study A2309 was to demonstrate comparable renal function between the everolimus groups and the Myfortic group. This review focused on the main safety endpoint of 12-month GFR and adverse events rates by system organ class, as well as, wound healing and edema. Please refer to the clinical review by Dr. Ergun Velidedeoglu and the safety statistical review by Dr. John Yap for more detailed safety analyses and evaluations.

3.2.1 Renal Function

The main safety endpoint for study A2309 was estimated GFR using the MDRD formula at 12 months post transplantation. Mean, standard deviation, median, and range of the estimated GFR at Month 12 are provided in Table 10. The everolimus 1.5 mg group had higher mean GFR than the Myfortic group (54.55 versus 52.18 mL/min/1.73m^2), with a p-value of 0.02. Even though the difference was statistically significant, a GFR difference of 2 mL/min/1.73m^2 was not considered as clinically meaningful by the clinical reviewer. The everolimus 3.0 mg group had slightly lower mean GFR than the Myfortic group (51.29 versus 52.18 mL/min/1.73m^2), but the difference was not statistically significant (p-value=0.92).

Note: Wilcoxon rank-sum test rather than t-test is used here, since the distribution of GFR was skewed and differed from a normal distribution.

For patients with missing 12-month GFR value, their GFR were imputed based on the pre-specified imputation method:

1. Patients who lost their graft were assigned a value zero for the 12-month GFR value.
2. Patients who died with a functioning graft had an imputed 12-month GFR value using the last-on-treatment-observation-carried- forward (LOCF) method.
3. Patients who had no GFR value past Day 315 and had neither died nor had graft loss would have an imputed 12-month GFR value using the LOCF method.

As sensitivity analyses, comparison of GFR at 12 months were conducted based on other imputation methods and similar results were found.

Table 10: Estimated GFR (MDRD) at 12 Month by Treatment Group (ITT Population)

Estimated GFR at Month 12 ($mL/min/1.73m^2$)	**everolimus 1.5 mg (N=277)**	**everolimus 3.0 mg (N=279)**	**Myfortic 1.44 g (N=277)**
Number without missing value *	275	278	277
Mean (± Standard Deviation)	54.55 (21.68)	51.29 (22.74)	52.18 (26.66)
Median (range)	55.0 (0, 140.9)	51.58 (0, 124.0)	49.7 (0, 366.4)
P-value ** (everolimus versus Myfortic)	0.02	0.92	N/A

* Three patients in the ITT population did not have any GFR value based on pre-specified imputation method, therefore were excluded from the analysis

** P-value for Wilcoxon rank-sum test

Note: The applicant's primary imputation using LOCF was based on last post-baseline on-treatment observation of GFR up to and including the scheduled Month 12 visit. Results from an analysis using the last post-baseline observation of GFR, including values observed during follow-up visits after discontinuation of study medication, during the 12-month study period resulted in similar results as to those presented in Table 10. Refer to the statistical safety review by Dr. John Yap for additional analyses of GFR.

To assess if different values of mean GFR at Month 12 were due to any baseline imbalances, baseline GFR and change of GFR from baseline to Month 12 were assessed (Table 11). Compared to the Myfortic group (8.71 $mL/min/1.73m^2$), the mean GFR at baseline was higher in both of the everolimus groups (9.32 and 9.15 $mL/min/1.73m^2$ respectively); however these differences were not statistically significant (p-value=0.84 and 0.38, respectively) as would be expected given the size of the study and randomization. Furthermore, the mean change of GFR from baseline to Month 12 was higher in the everolimus 1.5 mg group (45.26 $mL/min/1.73m^2$) than the Myfortic group (43.58 $mL/min/1.73m^2$), with a marginally significant p-value of 0.06. In the everolimus 3.0 mg group, the mean change of GFR from baseline to Month 12 was slightly lower (42.25 $mL/min/1.73m^2$) than in the Myfortic group, with a p-value of 0.95.

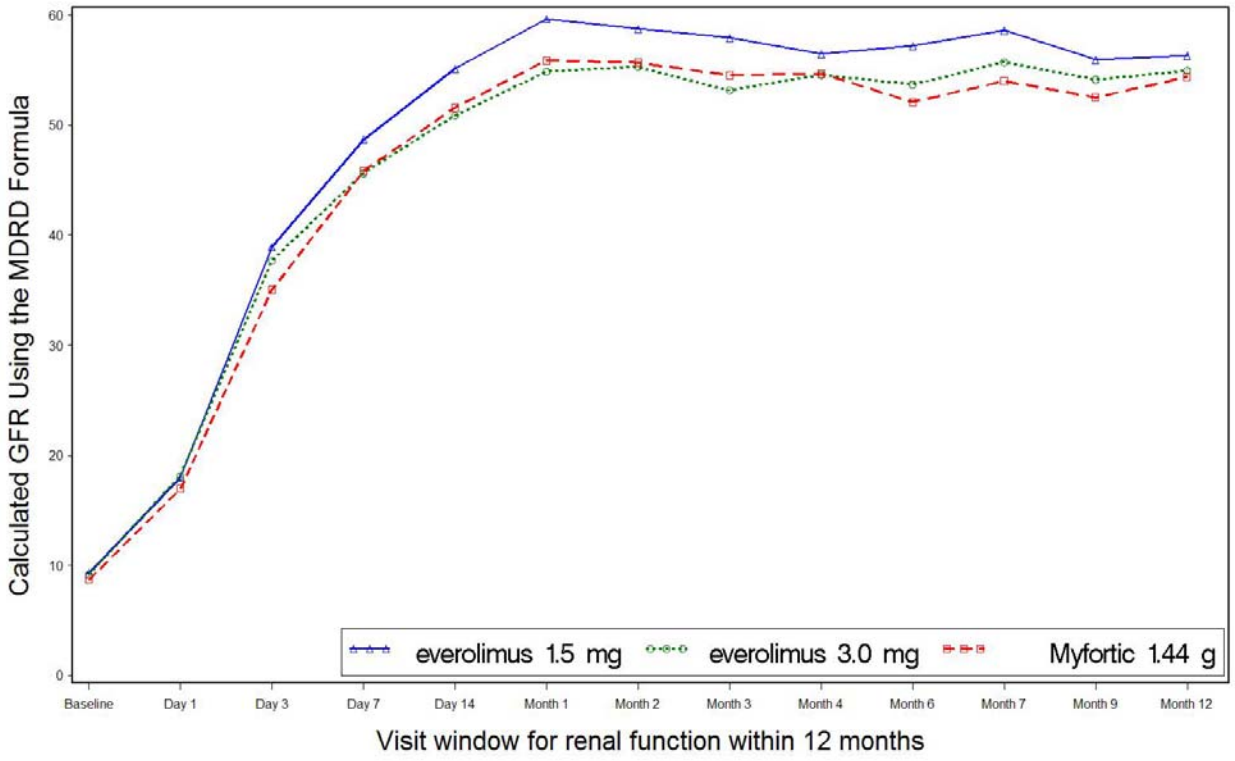
The mean estimated GFR values over different study visit windows within 12 months are summarized by treatment groups in Figure 6. The mean GFR was consistently higher, across the study period, in the everolimus 1.5 mg group compared to the mean in the Myfortic group. Additionally, the mean GFR was similar between the everolimus 3.0 group and the Myfortic group.

Table 11: Estimated GFR (MDRD) by Treatment Group (ITT Population - 12 Month Analysis)

Estimated GFR (mL/min/1.73m^2)	**everolimus 1.5 mg (N=277)**	**everolimus 3.0 mg (N=279)**	**Myfortic 1.44 g (N=277)**
GFR at baseline			
Number without missing value	267	270	271
Mean (± SD)	9.32 (7.90)	9.15 (5.42)	8.71 (4.53)
Median (range)	8.0 (2.8, 83.4)	7.85 (2.9, 49.8)	7.6 (2.1, 27.2)
P-value * (everolimus versus Myfortic)	0.84	0.38	N/A
Change of GFR from baseline to Month 12			
Number without missing value	266	270	271
Mean (± SD)	45.26 (23.08)	42.25 (23.47)	43.58 (27.68)
Median (range)	45.7 (-31.0, 130.1)	43.45 (-17.0, 115.9)	41.1 (-16.9, 359.1)
P-value * (everolimus versus Myfortic)	0.06	0.95	N/A

* P-value for Wilcoxon rank-sum test

Figure 6: Mean Estimated GFR (MDRD) over Time by Treatment Group (ITT Population)



3.2.2 Adverse Events

The incidence rates of adverse events or infections are shown by system organ class in Table 12. Almost all patients experienced at least one adverse event in all treatment groups. The most frequently affected organ classes were metabolism and nutrition disorders, and gastrointestinal disorders. More than 70% of patients per treatment group reported adverse events in these organ classes. Generally the incidence of adverse events by system organ class between treatment groups was similar. The system organ classes with the most notable differences with more events on everolimus 1.5 mg compared to Myfortic included general disorders and administration site conditions, metabolism and nutrition disorders, and reproductive system and breast disorders.

Note: Adverse events, unlike efficacy endpoints and GFR, were measured on treatment only (plus 8 days for non-serious and 30 days for serious adverse events after premature treatment discontinuation). Due to the imbalance in premature treatment discontinuation, with significantly more treatment discontinuations occurring in the two everolimus treatment groups compared to the Myfortic group, there was overall less follow-up time for adverse events in the two everolimus groups compared to the Myfortic group. Therefore, the comparison of adverse events across treatment groups is biased against the Myfortic group. This should be taken into consideration when interpreting the information in this section.

Table 12: Incidence Rates of Adverse Events/Infections by System Organ Class and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	everolimus 1.5mg (N=274)	everolimus 3.0mg (N=278)	Myfortic 1.44g (N=273)
Any system organ class	**271 (98.9)**	**276 (99.3)**	**270 (98.9)**
Blood and lymphatic system disorders	93 (33.9)	112 (40.3)	111 (40.7)
Cardiac disorders	43 (15.7)	39 (14.0)	43 (15.8)
Congenital, familial and genetic disorders	7 (2.6)	4 (1.4)	2 (0.7)
Ear and labyrinth disorders	13 (4.7)	4 (1.4)	14 (5.1)
Endocrine disorders	11 (4.0)	10 (3.6)	20 (7.3)
Eye disorders	29 (10.6)	22 (7.9)	28 (10.3)
Gastrointestinal disorders	196 (71.5)	209 (75.2)	207 (75.8)
General disorders and administration site conditions	182 (66.4)	187 (67.3)	161 (59.0)
Hepatobiliary disorders	7 (2.6)	8 (2.9)	8 (2.9)
Immune system disorders	14 (5.1)	9 (3.2)	11 (4.0)
Infections and infestations	170 (62.0)	180 (64.7)	188 (68.9)
Injury, poisoning and procedural complications	166 (60.6)	175 (62.9)	163 (59.7)
Investigations	137 (50.0)	120 (43.2)	134 (49.1)
Metabolism and nutrition disorders Musculoskeletal	222 (81.0)	233 (83.8)	199 (72.9)
and connective tissue disorders	112 (40.9)	107 (38.5)	105 (38.5)
Neoplasms benign, malignant and unspecified (incl cysts and polyps)	9 (3.3)	8 (2.9)	16 (5.9)
Nervous system disorders	92 (33.6)	96 (34.5)	109 (39.9)
Pregnancy, puerperium and perinatal conditions	1 (0.4)	0 (0.0)	0 (0.0)
Psychiatric disorders	90 (32.8)	76 (27.3)	72 (26.4)
Renal and urinary disorders	113 (41.2)	144 (51.8)	125 (45.8)
Reproductive system and breast disorders	50 (18.2)	52 (18.7)	23 (8.4)
Respiratory, thoracic and mediastinal disorders	87 (31.8)	109 (39.2)	93 (34.1)
Skin and subcutaneous tissue disorders	92 (33.6)	103 (37.1)	102 (37.4)
Social circumstances	0 (0.0)	1 (0.4)	1 (0.4)
Surgical and medical procedures	0 (0.0)	2 (0.7)	0 (0.0)
Vascular disorders	123 (44.9)	137 (49.3)	124 (45.4)

The subgroup analysis results of the incidence of adverse events or infections by gender are shown in Table 13. In general, more system organ classes were found with higher incidence in the everolimus groups than in the Myfortic group among female patients. The system organ classes with noticeably higher rate in the everolimus 1.5 mg group as compared to Myfortic among female patients included

- General disorders and administration site conditions
- Injury, poisoning and procedural complications
- Investigations
- Metabolism and nutrition disorders Musculoskeletal
- Reproductive system and breast disorders
- Vascular disorders

Among male patients, only the rate of reproductive system and breast disorders was

noticeably higher in the everolimus 1.5 mg group than in the Myfortic group.

Table 13: Incidence Rates of Adverse Events/Infections by System Organ Class, Gender, and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	Males Total =553			Females Total =272		
	EVR 1.5 mg (N=175)	EVR 3.0 mg (N=190)	Myfortic 1.44 g (N=188)	EVR 1.5 mg (N=99)	EVR 3.0 mg (N=88)	Myfortic 1.44 g (N=85)
Any system organ class	**173 (98.9)**	**188 (98.9)**	**186 (98.9)**	**98 (98.9)**	**88 (100)**	**84 (98.9)**
Blood and lymphatic	54 (30.9)	73 (38.4)	76 (40.4)	39 (39.4)	39 (44.3)	35 (41.2)
Cardiac disorders	22 (12.6)	33 (17.4)	29 (15.4)	21 (21.2)	6 (6.8)	14 (16.5)
Congenital/familial/genetic	5 (2.9)	4 (2.1)	2 (1.1)	2 (2.0)	0 (0.0)	0 (0.0)
Ear and labyrinth	8 (4.6)	2 (1.1)	9 (4.8)	5 (5.1)	2 (2.3)	5 (5.9)
Endocrine disorders	4 (2.3)	7 (3.7)	15 (8.0)	7 (7.1)	3 (3.4)	5 (5.9)
Eye disorders	16 (9.1)	18 (9.5)	20 (10.6)	13 (13.1)	4 (4.5)	8 (9.4)
Gastrointestinal	115 (65.7)	147 (77.4)	137 (72.9)	81 (81.8)	62 (70.5)	70 (82.4)
General disorders	110 (62.9)	133 (70.0)	110 (58.5)	72 (72.7)	54 (61.4)	51 (60.0)
Hepatobiliary	4 (2.3)	6 (3.2)	6 (3.2)	3 (3.0)	2 (2.3)	2 (2.4)
Immune system	6 (3.4)	6 (3.2)	6 (3.2)	14 (5.1)	9 (3.2)	11 (4.0)
Infections/infestations	104 (59.4)	117 (61.6)	125 (66.5)	66 (66.7)	63 (71.6)	63 (74.1)
Injury/poisoning	101 (57.7)	119 (62.6)	118 (62.8)	65 (65.7)	56 (63.6)	45 (52.9)
Investigations	84 (48.0)	85 (44.7)	101 (53.7)	53 (53.5)	35 (39.8)	33 (38.8)
Metabolism and nutrition	139 (79.4)	157 (82.6)	142 (75.5)	83 (83.8)	76 (86.4)	57 (67.1)
Musculoskeletal	72 (41.1)	76 (40.0)	71 (37.8)	40 (40.4)	31 (35.2)	34 (40.0)
Neoplasms	7 (4.0)	7 (3.7)	14 (7.4)	2 (2.0)	1 (1.1)	2 (2.4)
Nervous system	57 (32.6)	66 (34.7)	71 (37.8)	35 (35.4)	30 (34.1)	38 (44.7)
Pregnancy/puerperium	0 (0)	0 (0)	0 (0)	1 (1.0)	0 (0.0)	0 (0.0)
Psychiatric disorders	53 (30.3)	54 (28.4)	46 (24.5)	37 (37.4)	22 (25.0)	26 (30.6)
Renal and urinary	68 (38.9)	99 (52.1)	92 (48.9)	45 (45.5)	45 (51.1)	33 (38.8)
Reproductive system/breast	27 (15.4)	34 (17.9)	16 (8.5)	23 (23.2)	18 (20.5)	7 (8.2)
Respiratory/thoracic	54 (30.9)	78 (41.1)	65 (34.6)	33 (33.3)	31 (35.2)	28 (32.9)
Skin	65 (37.1)	70 (36.8)	70 (37.2)	27 (27.3)	33 (37.5)	32 (37.6)
Social circumstances	0 (0.0)	0 (0.0)	1 (0.5)	0 (0.0)	1 (1.1)	0 (0.0)
Surgical	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	1 (1.1)	0 (0.0)
Vascular disorders	74 (42.3)	92 (48.4)	93 (49.5)	49 (49.5)	45 (51.1)	31 (36.5)

Wound Healing Events

The antiproliferative effects of mTOR inhibitors, including everolimus, have previously been associated with delayed wound healing and fluid collections. The incidence of any wound healing related events was 6.2%, 14.4% and 5.1% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively within 12 months post transplantation (Table 14). As compared to the Myfortic group, the incidence was statistically significantly higher in the everolimus 3.0 mg group, with a difference of 9.3% (95% CI: 4.4, 14.2, p-value<0.001, Fisher's exact test). The incidence was also higher in the everolimus 1.5 mg group compared to Myfortic; however, the difference was not statistically significant (p-value=0.71).

Table 14: Summary of Wound Healing Related Event by Preferred Term and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	**everolimus 1.5mg (N=274)**	**everolimus 3.0mg (N=278)**	**Myfortic 1.44g (N=273)**
Any wound healing related event	**17 (6.2)**	**40 (14.4)**	**14 (5.1)**
Impaired healing	6 (2.2)	11 (4.0)	3 (1.1)
Wound dehiscence	5 (1.8)	14 (5.0)	4 (1.5)
Incisional hernia	5 (1.8)	7 (2.5)	3 (1.1)
Wound decomposition	1 (0.4)	0 (0)	0 (0)
Hernia	1 (0.4)	0 (0)	0 (0)
Abdominal wound dehiscence	0 (0)	5 (1.8)	2 (0.7)
Postoperative wound complication	0 (0)	2 (0.7)	1 (0.4)
Wound evisceration	0 (0)	1 (0.4)	0 (0)
Hernia obstructive	0 (0)	2 (0.7)	0 (0)
Incisional hernia, obstructive	0 (0)	0 (0)	1 (0.4)
Difference (everolimus - Myfortic) **95% CI** **P-value***	**1.1** **(-2.8, 5.0)** **p=0.71**	**9.3** **(4.4, 14.2)** **p<0.001**	N/A

Preferred terms were sorted by descending order of frequency in the everolimus 1.5 mg group
A patient with multiple occurrence of an event was counted only once in an event category
* P-value for Fisher's exact test

The subgroup analysis results of the incidence of wound healing related adverse event by gender are shown in Table 15. Among male patients, the incidence of any wound healing related events was 4.6%, 16.3% and 5.1% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. As compared to Myfortic, the incidence was statistically significantly higher in the everolimus 3.0 mg group (p-value<0.001), and the incidence was slightly lower in the everolimus 1.5 mg group with no statistical significance (p-value=1.0). Among female patients, the incidence of wound healing related event was higher in both everolimus groups than in the Myfortic group (9.1% and 10.2%, versus 5.9%), while the differences were not statistically significant (p-value=0.58 and 0.41, respectively).

Table 15: Summary of Wound Healing Related Event by Preferred Term, Gender, and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	Males Total =553			Females Total =272		
	everolimus 1.5 mg (N=175)	everolimus 3.0 mg (N=190)	Myfortic 1.44 g (N=188)	everolimus 1.5 mg (N=99)	everolimus 3.0 mg (N=88)	Myfortic 1.44 g (N=85)
Any wound healing related event	**8 (4.6)**	**31 (16.3)**	**9 (5.1)**	**9 (9.1)**	**9 (10.2)**	**5 (5.9)**
Impaired healing	3 (1.7)	7 (3.7)	1 (0.5)	3 (3.0)	4 (4.5)	2 (2.4)
Wound dehiscence	0 (0)	12 (6.3)	3 (1.6)	5 (5.1)	2 (2.3)	1 (1.2)
Incisional hernia	4 (2.3)	6 (3.2)	2 (1.1)	1 (1.0)	1 (1.1)	1 (1.2)
Wound decomposition	0 (0)	0 (0)	0 (0)	1 (1.0)	0 (0)	0 (0)
Hernia	1 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Abdominal wound dehiscence	0 (0)	2 (1.1)	1 (0.5)	0 (0)	3 (3.4)	1 (1.2)
Postoperative wound complication	0 (0)	2 (1.1)	1 (0.5)	0 (0)	0 (0)	0 (0)
Wound evisceration	0 (0)	1 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)
Hernia obstructive	0 (0)	2 (1.1)	0 (0)	0 (0)	0 (0)	0 (0)
Incisional hernia, obstructive	0 (0)	0 (0)	1 (0.5)	0 (0)	0 (0)	0 (0)
Difference (everolimus - Myfortic) **95% CI** **P-value***	**-0.5** **(-4.6, 4.1)** **p=1.0**	**9.2** **(5.5, 17.6)** **p<0.001**	N/A	**3.2** **(-2.8, 5.0)** **p=0.58**	**4.3** **(-3.7, 12.4)** **p=0.41**	N/A

Preferred terms were sorted by descending order of frequency in the everolimus 1.5 mg group
A patient with multiple occurrence of an event was counted only once in an event category
* P-value for Fisher's exact test

Edema

At month 12, there were notably more edema-related events in both everolimus groups compared to Myfortic (Table 14). The incidence of edema-related events was statistically significantly higher in both everolimus groups than in the Myfortic group (55.5% and 54.7% versus 45.1%, p-value was 0.02 and 0.03 respectively). The most common type of edema related event was peripheral edema, occurring 44.9%, 43.5% and 39.6% in everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively (p-value was 0.23 and 0.39 respectively).

Table 16: Summary of Edema-related Event by Preferred Term and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	**everolimus 1.5mg (N=274)**	**everolimus 3.0mg (N=278)**	**Myfortic 1.44g (N=273)**
Any edema related event	**152 (55.5)**	**152 (54.7)**	**123 (45.1)**
Oedema peripheral	123 (44.9)	121 (43.5)	108 (39.6)
Fluid overload	20 (7.3)	16 (5.8)	17 (6.2)
Oedema	20 (7.3)	16 (5.8)	14 (5.1)
Generalized oedema	6 (2.2)	6 (2.2)	3 (1.1)
Fluid retention	3 (1.1)	7 (2.5)	4 (1.5)
Pitting oedema	3 (1.1)	2 (0.7)	6 (2.2)
Gravitational oedema	1 (0.4)	0 (0)	0 (0)
Localized oedema	1 (0.4)	5 (1.8)	3 (1.1)
Oedema due to renal disease	1 (0.4)	0 (0)	0 (0)
Lymphoedema	0 (0)	1 (0.4)	0 (0)
Difference (everolimus - Myfortic) **95% CI** **P-value***	**10.4** **(2.1, 18.8)** **p=0.02**	**9.6** **(1.3, 17.9)** **p=0.03**	N/A

Preferred terms were sorted by descending order of frequency in the everolimus 1.5 mg group
A patient with multiple occurrence of an event was counted only once in an event category.
* P-value for Fisher's exact test

The subgroup analysis results of edema-related adverse event by gender are shown in Table 17. Among male patients, the incidence of edema-related events was higher in both everolimus groups than in the Myfortic group (49.7% and 55.8% versus 45.7%, p-value=0.46 and 0.06 respectively). Similarly, among female patients, the incidence was 65.7%, 52.3% and 43.5% in the everolimus 1.5 mg, 3.0 mg and Myfortic group respectively. As compared to Myfortic, the incidence was statistically significantly higher in the everolimus 1.5 mg group with p-value=0.003. The incidence was also numerically higher in the everolimus 3.0 mg group, but no statistically significant difference was found (p-value=0.29)

Table 17: Summary of Edema-related Event by Preferred Term, Gender, and Treatment Group (Safety Population - 12 Month Analysis)

Number of patients (%)	Males Total =553			Females Total =272		
	EVR 1.5 mg (N=175)	EVR 3.0 mg (N=190)	Myfortic 1.44 g (N=188)	EVR 1.5 mg (N=99)	EVR 3.0 mg (N=88)	Myfortic 1.44 g (N=85)
Any edema related event	**187 (49.7)**	**106 (55.8)**	**86 (45.7)**	**65 (65.7)**	**46 (52.3)**	**37 (43.5)**
Oedema peripheral	72 (41.1)	88 (46.3)	77 (41.0)	51 (51.5)	33 (37.5)	31 (36.5)
Fluid overload	10 (5.7)	8 (4.2)	13 (6.9)	10 (10.1)	8 (9.1)	4 (4.7)
Oedema	12 (6.9)	11 (5.8)	9 (4.8)	8 (8.1)	5 (5.7)	5 (5.9)
Generalized oedema	4 (2.3)	1 (0.5)	1 (0.5)	2 (2.0)	5 (5.7)	2 (2.4)
Fluid retention	2 (1.1)	5 (2.6)	1 (0.5)	1 (1.0)	2 (2.3)	3 (3.5)
Pitting oedema	1 (0.6)	2 (1.1)	6 (3.2)	2 (2.0)	0 (0)	0 (0)
Gravitational oedema	1 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Localized oedema	0 (0)	4 (2.1)	1 (0.5)	1 (1.0)	1 (1.1)	2 (2.4)
Oedema due to renal disease	1 (0.6)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Lymphoedema	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.1)	0 (0)
Difference (EVR - Myfortic) **95% CI P-value***	**4.0 (-6.3, 14.3) p=0.46**	**10.1 (0, 20.1) p=0.06**	N/A	**22.2 (8.0, 36.2) p=0.003**	**8.8 (-6.1, 23.6) p=0.29**	N/A

Preferred terms were sorted by descending order of frequency in the everolimus 1.5 mg group
A patient with multiple occurrence of an event was counted only once in an event category.
* P-value for Fisher's exact test

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Gender

Subgroup analysis of the primary efficacy endpoint by gender is presented in Table 18. Among male patients, the efficacy failure rate at 12 months post-transplantation was 28.4%, 21.5%, and 29.6%, in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. Compared to the Myfortic group, the everolimus 1.5 mg group had a slightly lower incidence of efficacy failure with risk difference of -1.2% (95% CI: -10.5, 8.1). The incidence was marginally significantly lower in the everolimus 3.0 mg group compared to the Myfortic group RD= -8.1% (-16.9, 0.6), p-value=0.08.

In contrast, the primary efficacy failure among female patients was more frequent in both everolimus groups than in the Myfortic groups. The incidence rate in the everolimus 1.5 mg, 3.0 mg and Myfortic groups was 19.0%, 22.7%, and 12.5% respectively. The difference between everolimus 1.5 mg and Myfortic was 6.5% (95% CI: -3.8, 16.8, p=0.24), and difference between everolimus 3.0 and Myfortic was 10.2% (95% CI: -0.9, 21.4, p=0.11). Additionally, a statistically significant interaction between treatment and gender (Breslow-Day test p-value=0.01) was indentified in the comparison of everolimus 3.0 mg to Myfortic. No statistically significant interaction between treatment and gender was found in the comparison between everolimus 1.5 mg to Myfortic (Breslow-Day test p-value=0.24). When interpreting these subgroup analysis results, one must take into account that multiple comparisons according to various subgroups were not adjusted.

Table 18: Primary Efficacy Endpoint Analysis by Gender and Treatment Group (ITT Population - 12 Month Analysis) [*]

Number of patients (%)	**Males Total =556**			**Females Total =276**		
	everolimus 1.5 mg (N=176)	**everolimus 3.0 mg (N=191)**	**Myfortic 1.44 g (N=189)**	**everolimus 1.5 mg (N=100)**	**everolimus 3.0 mg (N=88)**	**Myfortic 1.44 g (N=88)**
Efficacy Failure [*]	**50 (28.4)**	**41 (21.5)**	**56 (29.6)**	**19 (19.0)**	**20 (22.7)**	**11 (12.5)**
Treated BPAR	33 (18.8)	25 (13.1)	39 (20.6)	12 (12.0)	12 (13.6)	8 (9.1)
Graft Loss	7 (4.0)	7 (3.7)	7 (3.7)	5 (5.0)	6 (6.8)	2 (2.3)
Death	3 (1.7)	7 (3.7) [**]	6 (3.2)	4 (4.0)	3 (3.4)	0 (0)
Loss to follow-up	10 (5.7)	7 (3.7)	8 (4.2)	1 (1.0)	1 (1.1)	1 (1.1)
95% CI (everolimus –Myfortic)	(-10.5, 8.1)	(-16.9, 0.6)	N/A	(-3.8, 16.8)	(-0.9, 21.4)	N/A
P-value [***]	p=0.82	p=0.08		p=0.24	p=0.11	

[*] One subject's gender was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis. Also a statistically significant interaction between treatment and gender (Breslow-Day test p-value =0.01) was indentified in the comparison of everolimus 3.0 mg to Myfortic

[**] One patient who died 10 days after withdrew consent was included

[***] P-value for the Fisher's exact test

Subgroup analysis of the main secondary endpoint (graft loss, death, or loss to follow-up) by gender is presented in Table 19. The observed incidence of graft loss, death, or loss to follow-up was similar across all three treatment groups (12.5% and 10.5% in the everolimus groups versus 12.7% in the Myfortic group) in male patients. Among female patients, the rate of graft loss, death, or loss to follow-up was 11.0%, 12.5%, 5.7% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively (p=0.09, everolimus1.5 mg v. Myfortic; p=0.05 everolimus 3.0 mg v. Myfortic, Fisher's exact test). Additionally, a statistically significant interaction between treatment and gender (Breslow-Day test p-value=0.03) was indentified in the comparison of everolimus 3.0 mg to Myfortic. No

statistically significant interaction between treatment and gender was found in the comparison between everolimus 1.5 mg to Myfortic (Breslow-Day test p-value=0.11).

Table 19: Graft Loss, Death, or Loss to Follow-up by Gender and Treatment Group (ITT Population - 12 Month Analysis) [*]

	Males Total =556			**Females Total =276**		
	everolimus 1.5 mg (N=176)	**everolimus 3.0 mg (N=191)**	**Myfortic 1.44 g (N=189)**	**everolimus 1.5 mg (N=100)**	**everolimus 3.0 mg (N=88)**	**Myfortic 1.44 g (N=88)**
Graft Loss, Death or Loss to follow-up [^]	**21 (11.9)**	**20 (10.5)**	**23 (12.2)**	**10 (10.0)**	**11 (12.5)**	**3 (3.4)**
Graft Loss	7 (4.0)	7 (3.7)	7 (3.7)	5 (5.0)	6 (6.8)	2 (2.3)
Death	3 (1.7)	7 (3.7) [**]	6 (3.2)	4 (4.0)	3 (3.4)	0 (0)
Loss to follow-up	11 (6.3)	8 (4.2)	10 (5.3)	2 (2.0)	2 (2.3)	1 (1.1)
95% CI (everolimus – Myfortic) **P-value[***]**	(-6.9%, 6.5) p=1.0	(-8.1%, 4.7) p=0.63	N/A	(-0.4%, 13.6) p=0.09	(1.2 %, 17.0) p=0.05	N/A

[*] One subject's gender was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis
[**] One patient who died 10 days after withdrew consent was included
[***] P-value for the Fisher's exact test
[^] Statistically significant interaction between treatment and gender (Breslow-Day test p-value=0.03) was indentified in the comparison of everolimus 3.0 mg to Myfortic

Results for analysis of the primary safety endpoint (mean estimated GFR at month 12) by gender are shown in Table 20. Among both male and female patients, each of the everolimus groups was shown to have comparable estimated GFR as compared to the Myfortic group.

Table 20: Mean Calculated GFR (MDRD) at 12 Month by Gender and Treatment Group (ITT Population)

	Estimated GFR at Month 12 ($mL/min/1.73m^2$)	**everolimus 1.5 mg**	**everolimus 3.0 mg**	**Myfortic 1.44 g**
Males Total=556	Number without missing value	175	190	189
	Mean (± SD)	53.93 (21.51)	51.24 (22.47)	51.97 (29.34)
	Median (range)	53.5 (0, 134.1)	51.85 (0, 124.0)	49.1 (0, 366.4)
	P-value [*] (everolimus versus Myfortic)	0.07	0.87	N/A
Females Total=276	Number without missing value	100	88	88
	Mean (± SD)	55.65 (22.04)	51.39 (23.43)	52.61 (19.86)
	Median (range)	55.9 (0, 140.9)	51.43 (0, 111.3)	50.4 (0, 105.8)
	P-value [*] (everolimus versus Myfortic)	0.20	0.91	N/A

[*] P-value for Wilcoxon rank-sum test

Analysis results of premature discontinuation by gender are presented in Table 21. Among female patients, the incidence of premature treatment discontinuation in the everolimus 1.5 mg, 3.0 mg and Myfortic groups was 32.0% (32/100), 38.6% (34/88), and 15.9% (14/88) respectively, resulting in a p-value of 0.01 (everolimus 1.5 mg – Myfortic) and a p-value of 0.001 (everolimus 3.0 mg – Myfortic). Furthermore, in the everolimus 1.5 mg group, approximately 22% of the female patients prematurely discontinued treatment due to adverse events, which was statistically significantly higher (p-value= 0.004) than the Myfortic group (6.8%). Similarly, the incidence of premature treatment discontinuation due to adverse events in female patients in the everolimus 3.0 mg group was statistically significantly higher compared to the Myfortic group (21.6% versus 6.8%, with p-value=0.009). Additionally, female patients prematurely discontinued the study phase more frequently in the everolimus groups than the Myfortic group (14% and 11.8% versus 4.6%, p-value=0.04 and 0.16 respectively).

Differences in rates of premature treatment discontinuation were not observed among male patients in the study. Specifically, the incidence of premature treatment discontinuation among male patients in the everolimus 1.5 mg , 3.0 mg and Myfortic groups was 29.0% (51/176), 31.9% (61/191), and 24.3% (46/189) respectively (p-value=0.34 for everolimus 1.5 mg versus Myfortic and p-value=0.11 for everolimus 3.0 mg versus Myfortic). Study discontinuation, among male patients, was similar across all three groups (13.6% and 12.0% versus 12.7%).

Table 21: Premature Study Medication or Study Phase Discontinuation by Gender and Treatment Group (ITT Population - 12 Month Analysis) [*]

Number of patients (%)	**Males Total =556**			**Females Total =276**		
	everolimus 1.5 mg (N=176)	**everolimus 3.0 mg (N=191)**	**Myfortic 1.44 g (N=189)**	**everolimus 1.5 mg (N=100)**	**everolimus 3.0 mg (N=88)**	**Myfortic 1.44 g (N=88)**
Discontinued study medication	**51(29.0)**	**61(31.9)**	**46 (24.3)**	**32 (32.0)[#]**	**34 (38.6)[#]**	**14 (15.9)**
Adverse event(s)	28 (15.9)	38 (19.9)	20 (10.6)	22 (22.0)[#]	19 (21.6)[#]	6 (6.8)
Unsatisfactory therapeutic effect	8 (4.6)	9 (4.7)	9 (4.8)	3 (3.0)	5 (5.7)	4 (4.6)
Others	15 (8.5)	14 (7.3)	17 (9.0)	7 (7.0)	10 (11.4)	4 (4.5)
Discontinued study phase	**24 (13.6)**	**23 (12.0)**	**24 (12.7)**	**14 (14.0)[#]**	**10 (11.4)**	**4 (4.6)**
Subject withdrew consent	14 (8.0)	8 (4.2)	11 (5.8)	6 (6.0)	0 (0)	1 (1.1)
Death	3 (1.7)	6 (3.1)	6 (3.2)	4 (4.0)	3 (3.4)	0 (0)
Graft loss	6 (1.7)	5 (2.6)	6 (3.2)	3 (3.0)	5 (5.7)	1 (1.1)
Unknown	1 (0.6)	4 (2.1)	1 (0.5)	1 (1.0)	2 (2.3)	2 (2.3)

[*] One subject's gender was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis

[#] $p<0.05$ compared to Myfortic for Fisher's exact test

Note: Subgroup analyses of adverse events by gender were also conducted. More details can be found in Section 3.2.2

Age and Race

Subgroup analyses of the primary efficacy endpoint by recipient age and race are presented in Table 22 and Table 23, respectively. No significant differences were seen among treatments within the different age categories (i.e. <=50 and >50). Among Black patients, the observed incidence of efficacy failure was lower in both everolimus groups than in the Myfortic group (29.4% and 35.0% versus 38.5%); however, no statistically significant differences were found (p=0.47 and 0.82 respectively). Note that Black patients represent only 13.5% of the total study population; therefore, caution should be used when interpreting findings in this small subgroup.

Table 22: Primary Efficacy Endpoint Analysis by Recipient Age Category and Treatment Group (ITT Population - 12 Month Analysis) [*]

Number of patients (%)	**Recipient Age ≤ 50 Total =452**			**Recipient Age > 50 Total =380**		
	everolimus 1.5mg (N=156)	**everolimus 3.0mg (N=153)**	**Myfortic 1.44 g (N=143)**	**everolimus 1.5mg (N=120)**	**everolimus 3.0mg (N=126)**	**Myfortic 1.44 g (N=134)**
Efficacy Failure	**44 (28.2)**	**35 (22.9)**	**35 (24.5)**	**25 (20.8)**	**26 (20.6)**	**32 (23.9)**
Treated BPAR	30 (19.2)	23 (15.0)	26 (18.2)	15 (12.5)	14 (11.1)	21 (15.7)
Graft Loss	7 (4.5)	5 (3.3)	5 (3.5)	5 (4.2)	8 (6.4)	4 (3.0)
Death	4 (2.6)	4 (2.6)	1 (0.7)	3 (2.5)	6 (4.8) [**]	5 (3.7)
Loss to follow-up	8 (5.1)	6 (3.9)	4 (2.8)	3 (2.5)	2 (1.6)	5 (3.7)
95% CI (everolimus – Myfortic)	(-6.3, 13.7)	(-11.3, 8.1)	N/A	(-13.3, 7.2)	(-13.4, 6.9)	N/A
P-value [***]	p=0.51	p=0.79		p=0.65	p=0.55	

[*] One subject's age was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis

[**] One patient who died 10 days after withdrew consent was included

[***] P-value for the Fisher's exact test

Table 23: Primary Efficacy Endpoint Analysis by Race Category and Treatment Group (ITT Population - 12 Month Analysis) *

Number of patients (%)	Black Total =113			Non-Black Total =720		
	everolimus 1.5mg (N=34)	everolimus 3.0mg (N=40)	Myfortic 1.44 g (N=39)	everolimus 1.5mg (N=243)	everolimus 3.0mg (N=239)	Myfortic 1.44 g (N=238)
Efficacy Failure	**10 (29.4)**	**14 (35.0)**	**15 (38.5)**	**60 (24.7)**	**47 (19.7)**	**52 (21.9)**
Treated BPAR	7 (20.6)	9 (22.5)	12 (30.8)	38 (15.6)	28 (11.7)	35 (14.7)
Graft Loss	3 (8.8)	4 (10.0)	2 (5.1)	9 (3.7)	9 (3.8)	7 (2.9)
Death	0 (0)	2 (5.0)	3 (7.7)	7 (2.9)	8 (3.3) **	3 (1.3)
Loss to follow-up	2 (5.9)	1 (2.5)	1 (2.6)	10 (4.1)	7 (2.9)	8 (3.4)
95% CI (everolimus – Myfortic)	(-30.7, 12.6)	(-24.7, 17.8)	N/A	(-4.7, 10.4)	(-9.5, 5.1)	N/A
P-value ***	p=0.47	p=0.82		p=0.52	p=0.57	

* One subject was classified as non-black without specified actual race is included in the analysis
** One patient who died 10 days after withdrew consent was included
*** P-value for the Fisher's exact test

Subgroup analyses of the incidence of graft loss, death or loss to follow-up by recipient age and race are presented in Table 24 and Table 25, respectively. No significant differences were seen among treatments within the different age or race categories.

Table 24: Graft Loss, Death, or Loss to Follow-up by Age Category and Treatment Group (ITT Population - 12 Month Analysis)*

Number of patients (%)	Recipient Age ≤ 50 Total =452			Recipient Age > 50 Total =380		
	everolimus 1.5mg (N=156)	everolimus 3.0mg (N=153)	Myfortic 1.44 g (N=143)	everolimus 1.5mg (N=120)	everolimus 3.0mg (N=126)	Myfortic 1.44 g (N=134)
Graft Loss, Death or Loss to follow-up	**20 (12.8)**	**15 (9.8)**	**12 (8.4)**	**11 (9.2)**	**16 (12.7)**	**14 (10.5)**
Graft Loss	7 (4.5)	5 (3.3)	5 (3.5)	5 (4.2)	8 (6.4)	4 (3.0)
Death	4 (2.6)	4 (2.6)	1 (0.7)	3 (2.5)	6 (4.8) **	5 (3.7)
Loss to follow-up	10 (6.4)	7 (4.6)	6 (4.2)	3 (2.5)	3 (2.4)	5 (3.7)
95% CI (everolimus – Myfortic)	(-2.5, 11.4)	(-5.1, 8.0)	N/A	(-8.6, 6.0)	(-5.5, 10.0)	N/A
P-value ***	p=0.26	p=0.69		p=0.83	p=0.70	

* One subject's gender was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis
** One patient who died 10 days after withdrew consent was included
*** P-value for the Fisher's exact test

Table 25: Graft Loss, Death, or Loss to Follow-up by Race Category and Treatment Group (ITT Population - 12 Month Analysis) *

Number of patients (%)	**Black Total =113**			**Non-Black Total =720**		
	everolimus 1.5mg (N=34)	**everolimus 1.5mg (N=40)**	**everolimus 1.5mg (N=39)**	**everolimus 1.5mg (N=243)**	**everolimus 3.0mg (N=239)**	**Myfortic 1.44 g (N=238)**
Graft Loss, Death or Loss to follow-up	**5 (14.7)**	**7 (17.5)**	**6 (15.4)**	**27 (11.1)**	**24 (10.0)**	**20 (8.4)**
Graft Loss	3 (8.8)	4 (10.0)	2 (5.1)	9 (3.7)	9 (3.8)	7 (2.9)
Death	0 (0)	2 (5.0)	3 (7.7)	7 (2.9)	8 (3.3) **	3 (1.3)
Loss to follow-up	2 (5.9)	1 (2.5)	1 (2.6)	12 (4.9)	9 (3.8)	10 (4.2)
95% CI (everolimus – Myfortic)	(-17.1, 15.8)	(-14.2, 18.5)	N/A	(-2.6, 8.0)	(-3.6, 6.8)	N/A
P-value ***	p=1.0	p=1.0		p=0.36	p=0.64	

* One subject was classified as non-black without specified actual race is included in the analysis
** One patient who died 10 days after withdrew consent was included
*** P-value for the Fisher's exact test

The results of premature study discontinuation by recipient age and race category are presented in Table 26 and 27. In all the age or race subsets, the incidence of premature treatment discontinuation or study discontinuation was higher in both the everolimus groups than in the Myfortic group, the pattern was consistent among different age or race subgroups.

Table 26: Premature Study Medication or Study Phase Discontinuation by Age Category and Treatment Group (ITT Population - 12 Month Analysis) *

Number of patients (%)	**Age < 50 Total =452**			**Age ≥ 50 Total =380**		
	everolimus 1.5 mg (N=156)	**everolimus 3.0 mg (N=153)**	**Myfortic 1.44 g (N=143)**	**everolimus 1.5 mg (N=120)**	**everolimus 3.0 mg (N=126)**	**Myfortic 1.44 g (N=134)**
Discontinued study medication	**50 (32.1)**	**50 (32.7)**	**30 (21.0)**	**33 (27.5)**	**45 (35.7)**	**30 (22.4)**
Adverse event(s)	27 (17.3)	29 (19.0)	13 (9.1)	23 (19.2)	28 (22.2)	13 (9.7)
Unsatisfactory therapeutic effect	7 (4.5)	9 (5.9)	8 (5.6)	4 (3.3)	5 (4.0)	5 (3.7)
Others	16 (10.3)	12 (7.8)	9 (6.3)	6 (5.0)	12 (9.5)	12 (9.0)
Discontinued study phase	**27 (17.3)**	**17 (11.1)**	**12 (8.4)**	**11 (9.2)**	**16 (12.7)**	**16 (11.9)**
Subject withdrew consent	17 (10.9)	5 (3.3)	5 (3.5)	3 (2.5)	3 (2.4)	7 (5.2)
Death	4 (2.6)	4 (2.6)	1 (0.7)	3 (2.5)	5 (4.0)	5 (3.7)
Graft loss	5 (3.2)	3 (2.0)	4 (2.8)	4 (3.3)	7 (5.6)	3 (2.2)
Unknown	1 (0.6)	5 (3.3)	2 (1.4)	1 (0.8)	1 (0.8)	1 (0.8)

* One subject's age was unknown (in the everolimus 1.5 mg group) and is excluded from this analysis

Table 27: Premature Study Medication or Study Phase Discontinuation by Race Category and Treatment Group (ITT Population - 12 Month Analysis) *

Number of patients (%)	Black Total =113			Non-Black Total =720		
	everolimus 1.5 mg (N=34)	everolimus 3.0 mg (N=40)	everolimus 1.5 mg (N=39)	everolimus 1.5 mg (N=243)	everolimus 3.0 mg (N=239)	Myfortic 1.44 g (N=238)
Discontinued study medication	**11 (32.4)**	**18 (45.0)**	**11 (28.2)**	**72 (29.6)**	**77 (33.2)**	**49 (20.6)**
Adverse event(s)	7 (20.6)	9 (22.5)	4 (10.3)	43 (17.7)	48 (20.1)	22 (9.2)
Unsatisfactory therapeutic effect	1 (2.9)	4 (10.0)	4 (10.3)	10 (4.1)	10 (4.2)	9 (3.8)
Others	3 (8.8)	4 (10.0)	3 (8.8)	19 (7.8)	19 (7.9)	18 (7.6)
Discontinued study phase	**5 (14.7)**	**6 (15.0)**	**6 (15.4)**	**33 (13.6)**	**27 (11.3)**	**22 (9.2)**
Subject withdrew consent	2 (5.9)	1 (2.5)	2 (5.1)	18 (7.4)	7 (2.9)	10 (4.2)
Death	0 (0)	2 (5.0)	3 (7.7)	7 (2.9)	7 (2.9)	3 (1.3)
Graft loss	2 (5.9)	3 (7.5)	1 (2.6)	7 (2.9)	7 (2.9)	6 (2.5)
Unknown	1 (2.9)	0 (0)	0 (0)	1 (0.4)	6 (2.5)	3 (1.3)

* One subject was classified as non-black without specified actual race is included in the analysis

4.2 Other Special/Subgroup Populations

Subgroup analyses of the primary efficacy endpoint (composite consisting of treated BPAR, graft loss, death, or loss to follow-up) by baseline diabetic status are presented in Table 28. The incidence of efficacy failure was similar between the everolimus groups and the Myfortic group in all the subgroups, and no statistically significant difference were identified.

Note: The sample size of patients with baseline diabetes is small; therefore, caution should be used when interpreting findings in this small subgroup.

Table 28: Primary Efficacy Endpoint Analysis by Baseline Diabetic Status and Treatment Group (ITT Population - 12 Month Analysis)

Number of patients (%)	Diabetic at Baseline Total =175			Non-Diabetic at Baseline Total =658		
	everolimus 1.5 mg (N=58)	everolimus 3.0 mg (N=48)	Myfortic 1.44 g (N=69)	everolimus 1.5 mg (N=219)	everolimus 3.0 mg (N=231)	Myfortic 1.44 g (N=208)
Efficacy Failure	**13 (22.4)**	**15 (31.3)**	**17 (24.6)**	**57 (26.0)**	**46 (19.9)**	**50 (24.0)**
Treated BPAR	9 (15.5)	6 (12.5)	10 (14.5)	36 (16.4)	31 (13.3)	37 (17.8)
Graft Loss	4 (6.9)	6 (12.5)	1 (1.5)	8 (3.7)	7 (3.0)	8 (3.9)
Death	2 (3.5)	5 (10.4)	5 (7.3)	5 (2.3)	4 (1.7)	1 (0.5)
Loss to follow-up	0 (0)	1 (2.1)	2 (2.9)	12 (5.5)	7 (3.0)	7 (3.4)
95% CI (everolimus – Myfortic)	(-17.0,12.6)	(-10.0, 23.2)	N/A	(-6.2, 10.2)	(-11.9, 3.6)	N/A
P-value *	p=0.84	p=0.53		p=0.66	p=0.30	

* P-value for the Fisher's exact test

Subgroup analyses of incidence of graft loss or death by patient's GFR at Month 1 are presented in Table 29. More than 13% of patients who had GFR less than 40 (mL/min/1.73m^2) at Month 1 were reported for graft loss or death within 12 months post kidney transplantation. The incidence of graft loss or death was higher in both everolimus groups than in the Myfortic group (14.3% and 15.0% versus 10.4%), but no statistically significant difference was shown. In contrast, among patients with GFR greater than 40 at Month 1, only approximately 2% had graft loss or death within 12 months. The incidence was similar across all three treatment groups. Note that more than 27% of patients with missing GFR value at Month 1 experienced graft loss or death in 12 months (26.9%, 29.6% and 26.9% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups). Since these patients represent less than 10% of the total ITT population, caution should be used when interpreting findings in this subgroup.

Note: Subgroup analyses in Table 29 by GFR at Month 1 are based on post-randomization variable and therefore conclusions regarding treatment and outcome in these subgroups should be interpreted with caution. The purpose of these subgroup analyses is to provide information to the clinical reviewer about the relationship between the treatment effect on graft loss or death at month 12 and the possible treatment effect on GFR at month 1.

Table 29: Graft Loss or Death Analysis by GFR at Month 1 (ITT Population - 12 Month Analysis)

Number of patients (%)	**GFR at Month 1 <40** (mL/min/1.73m^2) **Total =150**			**GFR at Month 1 ≥40** (mL/min/1.73m^2) **Total =604**		
	everolimus 1.5 mg (N=42)	**everolimus 3.0 mg (N=60)**	**Myfortic 1.44 g (N=48)**	**everolimus 1.5 mg (N=209)**	**everolimus 3.0 mg (N=192)**	**Myfortic 1.44 g (N=203)**
Graft Loss or Death Graft Loss Death	**6 (14.3)** 5 (11.9) 1 (2.4)	**9 (15.0)** 7 (11.7) 2 (3.3)	**5 (10.4)** 4 (8.3) 1 (2.1)	**5 (2.4)** 1 (0.5) 4 (1.9)	**4 (2.1)** 2 (1.0) 3 (1.6)	**3 (1.5)** 1 (0.5) 2 (1.0)
95% CI (everolimus – Myfortic) **P-value** *	(-22.4, 40.4) p=0.75	(-17.0, 37.1) p=0.57	N/A	(-21.9, 45.9) p=0.72	(-28.3 , 45.7) p=0.72	N/A
	GFR missing at Month 1 Total =79					
	everolimus 1.5 mg (N=26)	**everolimus 3.0 mg (N=27)**	**Myfortic 1.44 g (N=26)**			
Graft Loss or Death Graft Loss Death	**7 (26.9)** 6 (23.1) 2 (7.7)	**8 (29.6)** 4 (14.8) 4 (14.8)	**7 (26.9)** 4 (15.4) 3 (11.5)			
95% CI (everolimus – Myfortic) **P-value** *	(-30.6, 30.6) p=1.0	(-26.5, 33.2) p=1.0	N/A			

* P-value for the Fisher's exact test

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The primary efficacy endpoint of study A2309 was a composite endpoint consisting of treated BPAR episode, graft loss, death, or loss to follow-up at 12 months post-transplant. Using the protocol-defined Hochberg's procedure for multiple comparison adjustment, the study demonstrated that both of the everolimus treatment regimens were non-inferior to the Myfortic treatment regimen in preventing the incidence of efficacy failure at 12 months. Rates of graft loss, death and loss to follow-up were deemed acceptable in both everolimus treatment regimens although the rates were numerically higher than that observed in the Myfortic group. Similar results between the everolimus and Myfortic groups were shown for other secondary efficacy endpoints.

Note: There is no justified non-inferiority margin for these endpoints.

The primary safety endpoint of study A2309 was estimated GFR using the MDRD formula at 12 months following the kidney transplantation. Similar values of estimated GFR at month 12 were achieved in each of the everolimus regimens compared to the Myfortic regimen.

Premature treatment discontinuation, primarily due to adverse events, was frequent and statistically significantly higher in each of the everolimus groups, as compared to the Myfortic group. The imbalanced incidence of treatment discontinuation should be of concern when interpreting the safety and efficacy outcomes of Study A2309 (more details could be found in Table 2, Figure 2, Section 3.1.2, 3.1.3, and 3.2.2).

Both the patient and the investigator were unblinded to the treatment regimen a patient received, because of the open-label design of study A2309. This should be taken into consideration in the interpretation of the study results, since unblinded study is more subject to bias. This is of particular concern given the observed higher rates of premature treatment discontinuation in both everolimus groups, which may be related to the un-blinded nature of the study

Subgroup analyses showed that efficacy results of Study A2309 were not consistent across gender. There was a significant interaction noted between the everolimus 3.0 mg group and Myfortic by gender. The efficacy failure rate was lower in both everolimus groups than in the Myfortic group in male patients. Among female patients, the efficacy failure rate was higher in both everolimus groups than Myfortic. Furthermore, for female patients, incidence of premature treatment discontinuation was significantly higher in each of the everolimus groups than in the Myfortic group. Most concerning, however, is the higher rates of graft loss and death in the everolimus subjects compared to Myfortic. Study A2309 may not provide adequate information to determine a safe and efficacious everolimus regimen for females (more details could be found in Table 18, Table 19 and Table 21, Section 4.1).

5.2 Conclusions and Recommendations

Based on protocol specified and justified 10% non-inferiority margin, study A2309 demonstrated that both everolimus treatment regimens were non-inferior to the Myfortic treatment regimen at 12 months in the incidence rate of efficacy failure (composite of treated BPAR, graft loss, death or loss to follow-up). The incidence rate of efficacy failure was 25.3%, 21.9% and 24.2% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. The difference between everolimus 1.5 mg and Myfortic was 1.1%, with 95% confidence interval (-6.1, 8.3). The difference between everolimus 3.0 mg and Myfortic was -2.3%, with 95% confidence interval (-9.3, 4.7). Additionally, the 12-month incidence of graft loss, death and loss to follow-up was similar between both everolimus groups and the Myfortic group, although numerically these events were more frequent in the everolimus groups compared to the Myfortic group.

As compared to the Myfortic treatment regimen, both everolimus treatment regimens were demonstrated to have similar renal function measured as estimated mean glomerular filtration rate (GFR) at 12 months post-transplantation.

There was a disproportionate rate of premature treatment discontinuation within 12 months in study A2309, driven by higher rates of adverse events in both everolimus groups compared to Myfortic. The incidence rate of premature treatment discontinuation was 30.0%, 34.7% and 21.7% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively. The rates were statistically significantly higher in both everolimus groups (p=0.03, everolimus 1.5 mg, p=0.001 everolimus 3.0 mg) compared to the Myfortic group. More patients in both of the everolimus groups prematurely discontinued study treatment and were subsequently switched to alternate therapy than in the Myfortic group, which may bias the interpretation of the study safety and efficacy results. Confidence intervals obtained from a sensitivity analysis including premature treatment discontinuations as failures in the primary efficacy endpoint could not rule out that everolimus was no more than 10% worse than Myfortic.

The incidence of edema-related events was statistically significantly higher in both everolimus groups than in the Myfortic group (55.5% and 54.7% versus 45.1%, p-value was 0.02 and 0.03 respectively). Additionally, the incidence of any wound healing related events was 6.2%, 14.4% and 5.1% in the everolimus 1.5 mg, 3.0 mg and Myfortic groups respectively within 12 months post transplantation ($p<0.001$ everolimus 3.0 mg versus Myfortic, p=0.71 everolimus 1.5 mg versus Myfortic).

Analysis by gender revealed that among female patients, rates of premature treatment discontinuation, primary efficacy failure and graft loss and death were considerably higher in both everolimus groups compared to the Myfortic group. No differences were seen in the subgroup analyses by age (< 50 years and ≥ 50 years) or by race (Black versus non-Black), although only about 14% of patients were Black, the others were Caucasian, Asian and other races.

Appendix 1: Justification of the non-inferiority margin

The applicant did not submit a justification for their non-inferiority margin in this NDA submission; instead they submitted the justification to their IND, IND 52,003 SN919/SDN981, on 6/26/09. This appendix contains the statistical review by Dr. LaRee Tracy of this IND submission.

Background:

This submission contains the sponsor's justification for the 10% non-inferiority (NI) margin for the Phase III study A2309. Study A2309 is complete and results were submitted to NDA 21-560 on 6/30/09 in support of a Class II re-submission to the pending application for Certican (everolimus). ***Note: This justification was not included in the complete response submission.***

The objective of study A2309 was to demonstrate that at least one everolimus treatment regimen including reduced dose Neoral is non-inferior, based on a 10% margin, to a Myfortic regimen given in combination with standard dose Neoral regimen. ***The primary endpoint was efficacy failure defined as incidence of the composite endpoint consisting of a treated biopsy proven acute rejection (BPAR) episode, graft loss (GL), death (D) or lost to follow-up (LTF) at 12 months.*** Treated BPAR was defined as a biopsy graded IA, IB, IIA, IIB, or III and which is treated with anti-rejection therapy. GL was defined as requiring dialysis or re-transplant.

Treatment arms included:

Experimental 1:	B+CS+CsA(r) + everolimus 1.5 mg
Experimental 2:	B+CS+CsA(r) + everolimus 3.0 mg
Control:	B+CS+CsA(s) + EC-MPS

Where B=basiliximab (Simulect®), CS=corticosteroids, CsA=cyclosporine A (Neoral®), EC-MPS=enteric-coated mycophenolate sodium (Myfortic®). Neoral was used with a reduced exposure (r) in the experimental arms and with the standard exposure (s) in the control arms. In the sponsor's justification, the active metabolite mycophenolate acid (MPA) of both pro-drugs EC-MPS and MMF (mycophenolate mofetil) are used as their synonyms.

The NI margin should not be larger than the amount of efficacy the control arm has over the putative placebo. To determine an appropriate NI margin, the 'control effect' needs to be determined. This is done via assessing the difference or ratio between the putative placebo and the control arm, i.e. B+CS+CsA(r) v. B+CS+CsA(s)+MPA using data from previously conducted clinical trials.

Experimental:	*B+CS+CsA(r) + everolimus*
Putative Placebo:	*B+CS+CsA(r)*
Control:	*B+CS+CsA(s) + MPA*

Reviewer's Comment: Ideally data to determine the control effect should come from randomized clinical trials (RCTs) in which the active control regimen was studied concurrently to the putative placebo.

Findings from Sponsor's Literature Search/Review

The sponsor performed a literature search to identify all relevant RCTs in *de novo* kidney transplantation, excluding trials conducted in special populations (e.g. pediatrics, DGF only, non-heart beating donor). The endpoint of interest was BPAR (and/or composite endpoint including BPAR) at 6- or 12-months post-transplant. Only studies including treatments with drugs or drug classes: CS, CNI, anti-IL2, MPA, mTOR-inhibitor, and FTY720 were included. The search yielded **51** relevant clinical trials published between 1996 and 2008 of which the smallest trial enrolled 67 patients and the largest enrolled 1,589 patients.

-No studies were identified that concurrently compared B+CS+CsA+MPA to B+CS+CsA in renal transplantation. ***Therefore, a meta-analysis of studies to estimate the control effect over putative placebo could not be performed.***

-No studies were identified that evaluated the use of B+CS+CsA (the putative placebo) in renal transplantation. ***Therefore, directly estimating the failure rate of this regimen to the control arm (B+CS+CsA+MPA) can not be done.***

-Many studies did not include a 12-month analysis. ***As such, the 6-month analysis served as a surrogate assuming no difference between 6-and 12-month time points.***

Reviewer's Comment: This assumption is sometimes true in that the majority of BPARs occur within the first 6-months post-transplant in kidney transplantation.

-Many studies did not report the incidence of the composite endpoint used in study A2309. ***Therefore the BPAR endpoint was used as surrogate while assuming no major differences in treatment effects between BPAR alone versus the composite endpoint***. The sponsor is basing this assumption on the observation that patients with a GL will likely also have a BPAR event prior to the GL; LTF occurs infrequently in quality renal transplantation studies; the incidence of patient death in renal transplantation studies is low or those who die experience have a BPAR event prior to death.

Reviewer's Comment: The assumptions above, i.e. BPAR precedes GL or D, low LTF in renal transplantation and there are few deaths that occur without GL, are generally true in kidney transplantation. The incidence of the composite endpoint in renal transplantation studies is generally driven by the rate of BPAR so in the absence of data on patient death, graft loss or loss to follow-up, this approach is acceptable to estimate the treatment failure rate.

-There was considerable variation in the treatment regimens (i.e. drugs used and doses) in the studies identified. Therefore, accounting for all combinations of regimens in the statistical models would have lead to complex models that will prevent useful estimation. For simplicity, the sponsor categorized each calcineurin-inhibitor (CNI) into 'reduced' or 'standard' drug exposure categories. All other dose-effects of the other treatment compounds contributed to the random study effect (δ_i) of the statistical model.

Sponsor's Assumptions for Modeling

-The efficacy of basiliximab and daclizumab (monoclonal antibodies with activity against the α chain of the interleukin-2 receptor, IL-2α) are assumed equivalent.

Reviewer's Comment: No RCTs have been performed to date directly comparing the effect of basiliximab to daclizumab. While the two monoclonal antibodies are not identical, they both target the same epitope on the same component of the same receptor and have the same mechanism of action (blocking, not lysing). The literature (including a Cochrane review on the topic) suggests that the antibodies likely have equivalent or near-equivalent contributions.

-The efficacy of MMF and EC-MPS are assumed to be equivalent.

Justification: MPA is an active metabolite of both pro-drugs EC-MPS and MMF. Two clinical trials investigated the bioequivalence (BE) between single and multiple doses of EC-MPS and MMF in renal transplant. BE was estimated based on a 90% CI of the EC-MPS:MMF ratio for the AUC and the C_{max} within the pre-specified limits (0.80, 1.25) based on clinical pharmacology requirements. As discussed below, the sponsor provides results from additional mixed effects models that treat EC-MPS and MMF as separate factors versus MPA, the reference factor.

-Efficacy of the mTOR-inhibitors, everolimus and sirolimus, are assumed to be equivalent.

Reviewer's Comment: This has not been established in a RCT.

- To extrapolate to B+CS+CsA(r), the effects of the individual components of a regimen were assumed to be additive on a log-odds (logit) scale.

Reviewer's Comment: This assumes the addition of one drug to another has an additive effect rather than a multiplicative one. This is a major assumption, which if incorrect may serious bias the results from the models.

Description of Modeling Approach

Given the lack of any RCTs evaluating the efficacy of the putative placebo in renal transplant patients, the sponsor used mixed effects modeling (MEM) to estimate the contribution of each of the immunosuppressant drugs to the combination therapy event rate. These models assume additive drug effects in a combination therapy in the log-odds scale as noted above. This model was used to estimate the combined effect of the three immunosuppressant drugs (B+CS+CsA(r)) for the putative placebo group and for the four (B+CS+CsA(s)+MPA) for the control group.

The mathematical model is as follows:

Let

$Y_{ij} \sim Binomial(N_{ij,}\pi_{ij})$ where Y_{ij} is the number of events in study i and treatment arm j

$N_{ij,}$ is the number of patients and π_{ij} is the event rate in study i and treatment arm j.

The log-odds of the event rate is related to the study effects based on a linear model with the random study effect:

$\text{logit}(\pi_{ij}) = \mu + \delta_i + x_{ij}\beta$ where the intercept and effects of the immunosuppressant drugs and covariates are μ and β respectively, the random effect $\delta_i \sim N(0,\sigma^2)$, vector x_{ij} includes indicators for presence or absences of each of the immunosuppressant drugs and covarates

The model was fit using the maximum-likelihood method (PROC NLMIXED in SAS). Plots of predicted v. observed event rates were used to assess model fits.

Reviewer's Comment: The sponsor's models assume an additive nature of each treatment to the overall regimen, which is an un-testable assumption given that there are no actual data to compare the model estimates.

Results from MEM:

The sponsor concluded that the estimated failure rate of the control group (B+CS+CsA(s)+MPA) is 18.8%, 95% CI (16.4, 21.4) and that for the putative placebo (B+CS+CsA(r)) is 43.5%, 95% CI (37.5, 49.4) and an estimated difference in event rates (B+CS+CsA(r)- B+CS+CsA(s)+MPA) of ***24.6% (18.9, 30.2)*** (Table 1). The sponsor further concluded, based on an estimated lower bound of 18.9% and if a 10% NI margin is chosen (as is in the protocol), 47% of the control effect can be preserved. ***Note: The estimate is the probability of the primary endpoint occurring.***

Reviewer's Comments:

1) The method used by the sponsor to derive the amount of preservation of the effect size is an approach among several approaches.

2) A 10% NI margin for the 12-month composite endpoint is acceptable given that the estimated lower bound around the estimated difference between the active control and the putative placebo is 18.9%. Although the lower bound is 18.9%, a margin larger than 10%; however, would be considered too large from a clinical perspective.

Table 1: Results from MEM for the event rate of the primary endpoint (12-mo composite)

Regimen	Estimate	Standard Error	95% Interval
Anti-IL2+CS+CsA(s)+MPA	0.1888	0.01253	(0.164,0.214)
Anti-IL2+CS+CsA(r)+MPA	0.2411	0.01757	(0.206, 0.276)
Anti-IL2+CS+CsA(s)	0.3602	0.02464	(0.311, 0.410)
Anti-IL2+CS+CsA(r)	0.4346	0.02966	(0.375, 0.494)
Anti-IL2+CS+CsA(r) - Anti-IL2+CS+CsA(s)+MPA	**0.2458**	**0.02807**	**(18.9, 30.2)**

Table re-created and modified from sponsor's table 3-2 in submission
N=51 studies, estimated CIs are based on the t-distribution (PROC NLMIXED)
CS=corticosteroids, CsA=cyclosporine A (Neoral®), (r)=reduced dose CsA, (s)=standard dose CsA, mycophenolate acid (MPA)

Sponsor's Sensitivity Analyses

The sponsor also performed sensitivity analyses using the same modeling approach as described above while excluding studies that included a) mTOR-inhibitors, b) mTOR-inhibitors and FTY720, c) mTOR-inhibitors, FTY720 and azathrioprine (AZA), and d) mTOR-inhibitors, FTY720, AZA and tacrolimus (Tac). Results from these analyses are presented in Table 2. Each row represents the estimated difference between the putative placebo and the active control regimen minus the respective immunosuppressant(s). The analysis excluding all studies with an mTOR-inhibitor, FTY720, AZA, or tacrolimus resulted in a lower bound around the difference in event rates of 18.7% (based on MEM of 14 studies, last row of Table 2). This was the smallest effect size estimated among these sensitivity analyses.

Table 2: Results from Sensitivity Analyses of MEM

Comparison	#Studies	Estimate	Std Error	95% Interval
Anti-IL2+CS+CsA(r) - Anti-IL2+CS+CsA(s)+MPA				
Without mTOR-Inhibitor	37	0.3193	0.03201	(0.254, 0.384)
Without mTOR-Inhibitor, FTY720	33	0.2905	0.04163	(0.206, 0.375)
Without mTOR-Inhibitor, FTY720, AZA	24	0.3397	0.05800	(0.235, 0.445)
Without mTOR-Inhibitor, FTY720, AZA, Tac	14	0.3066	0.05549	(0.187, 0.426)

Table re-created using data in sponsor's table 3-3
CS=corticosteroids, CsA=cyclosporine A (Neoral®), (r)=reduced dose CsA, (s)=standard dose CsA, mycophenolate acid (MPA), AZA=azathioprine, Tac=tacrolimus

The sensitivity analyses are consistent with the results from the main analysis, which estimated a lower bound around the estimated difference in event rates of 18.9%.

Comparison of EC-MPS and MMF

As noted above, the BE between EC-MPS and MMF was previously evaluated. To evaluate these different treatments, the sponsor developed models that included EC-MPS and MMF as separate factors versus the common factor MPA and then estimated the difference between the two compounds. These estimates and comparisons using MEMs are shown below in Table 3 suggesting little difference between the two compounds (last two rows of Table 3) with standard and reduced dose Neoral.

Table 3: Model Estimates for Treatment Regimens for EC-MPS and MMF on Primary Endpoint (12-month composite)

Regimen	Estimate	Std Error	95% Interval
Anti-IL2+CS+CsA(s)+EC-MPS	0.1940	0.02127	(0.151, 0.237)
Anti-IL2+CS+CsA(s)+MMF	0.1874	0.01322	(0.161, 0.214)
Anti-IL2+CS+CsA(r)+EC-MPS	0.2476	0.02768	(0.192, 0.303)
Anti-IL2+CS+CsA(r)+MMF	0.2397	0.01806	(0.203, 0.276)
Difference			
Anti-IL2+CS+CsA(**s**)+EC-**MPS**- Anti-IL2+CS+CsA**(s)**+**MMF**	0.0066	0.02145	(-0.036, 0.050)
Anti-IL2+CS+CsA(**r**)+EC-**MPS**- Anti-IL2+CS+CsA**(r)**+**MMF**	0.0079	0.02599	(-0.043, 0.059)

Table re-created from table 6-1 of submission

CS=corticosteroids, CsA=cyclosporine A (Neoral®), (r)=reduced dose CsA, (s)=standard dose CsA, mycophenolate acid (MPA), AZA=azathioprine, Tac=tacrolimus

Constancy Assumption

A key component of NI margin justification is the constancy assumption, which assumes that the clinical trials used to justify the NI margin are similar in design, patient population, medical care/intervention, and clinical setting to that of the planned RCT. This is an important assumption that must hold true given the biases that ensue when using results from trials that do not resemble the planned trial.

To test the constancy assumption, the sponsor fit MEMs by year of publication to assess for changes in estimate as a function of time. The sponsor notes that use of *reduced* CsA before 2005 was sparse, therefore the estimation of the control effect over time would be dominated by the estimation of the antil-IL2 and CS over time; whereas the estimation of CsA(r) is based on data from 2005-2008. Therefore, instead of comparing the difference between putative placebo and control over time, the sponsor assessed the control estimates over time. The graph produced (not shown here) does not suggest any major fluctuations in estimated failure rate in the control group between 1996 and 2008. The estimated between-study standard deviation from the random study effect was 0.2204 on a log-odds scale.

Reviewer's Comment: This approach to assess constancy is acceptable when accepting all the stated assumptions made for the mixed effects models and while assuming all other parameters are constant. The shift towards reduced CsA after

2005 was driven more by the need to reduce associated renal toxicity and less with the primary endpoint of BPAR. Therefore, though the sponsor's assessment of constancy suggests a lack of variation across time, this is only with respect to one endpoint.

Assessment of Model Fit

Lastly, to assess the models' assumption that the effects of the treatments are additive on a log-odds scale, the sponsor generated a plot of predicted v. observed rates of all treatment in all studies. This plot (not presented here) suggests a general linear association between proportion of events and predicted proportion of events.

Reviewer's Comment: This approach could be strengthened by plotting model residuals against predicted estimates.

Reviewer's MEMs of Available Studies

The following is a summary of the reviewer's analysis of the 51 clinical trials identified by the sponsor.

Note: CsA0 = CsA(standard), CsA1 = CsA(reduced),
Tacl = Tac(standard), TacO = Tac(reduced)

All 51 studies

The NLMIXED Procedure

Specifications

Data Set	WORK.TWO
Dependent Variable	MX_K
Distribution for Dependent Variable	Binomial
Random Effects	delta
Distribution for Random Effects	Normal
Subject Variable	NR
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	123
Observations Not Used	0
Total Observations	123
Subjects	51
Max Obs Per Subject	4
Parameters	14
Quadrature Points	1

All 51 studies

The NLMIXED Procedure

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper

logodds (CS +antiil2 + CsA0(s) + MMF/MPA)	-1.4581	0.08182	50	-17.82	<.0001	0.05	-1.6224	-1.2937
logodds (CS +antiil2 + CsA1(r) + MMF/MPA)	-1.1468	0.09602	50	-11.94	<.0001	0.05	-1.3396	-0.9539
logodds (CS +antiil2 + CsA0(s))	-0.5744	0.1069	50	-5.37	<.0001	0.05	-0.7891	-0.3596
logodds (CS +antiil2 + CsA1(r))	-0.2631	0.1207	50	-2.18	0.0340	0.05	-0.5055	-0.02064
prob (CS)	0.8599	0.02790	50	30.82	<.0001	0.05	0.8039	0.9160
prob (CS +antiil2 + CsA0(s) + MMF/MPA)	0.1888	0.01253	50	15.07	<.0001	0.05	0.1636	0.2139
prob (CS +antiil2 + CsA1(r) + MMF/MPA)	0.2411	0.01757	50	13.72	<.0001	0.05	0.2058	0.2764
prob (CS +antiil2 + CsA0(s))	0.3602	0.02464	50	14.62	<.0001	0.05	0.3107	0.4097
prob (CS +antiil2 + CsA1(r))	0.4346	0.02966	50	14.65	<.0001	0.05	0.3750	0.4942
prob (CS +antiil2+CsA1(r))- (CS+antiil2+CsA0(s)+MPA)	0.2458	0.02807	50	8.76	<.0001	0.05	0.1895	0.3022
prob (CS +antiil2+CsA1(s))- (CS+antiil2+CsA0(s)+MPA)	0.1715	0.01968	50	8.71	<.0001	0.05	0.1319	0.2110

Reviewer's Comment: The yellow highlighted rows are the estimates for the control group and the putative placebo group based on the reviewer's analysis. The sponsor did not provide the dataset used to derive their estimates and therefore the reviewer utilized a modified version of a dataset provided to the Division for a similar justification, which included most of the identified studies (the reviewer added the missing studies). The green highlighted row is the estimated difference between putative placebo and the active control. These results are the same as presented by the sponsor.

Additional estimates of effect size

The above justification for the NI margin is that proposed by the sponsor and leads to an effect size large enough to support their proposed 10% margin. This reviewer proposes an additional three analyses to assess the robustness of the sponsor's conclusions.

1) In study A2309, the test drug, everolimus, is given with a reduced dose of cyclosporine. Therefore, the justification of the NI margin should be based on the effect of MPA, which is replaced by everolimus, as well as, a portion of the effect of cyclosporine (CNI) in the control arm. MPA is known to interact with cyclosporine so that higher doses are needed to obtain similar effects as would be needed with lower doses when given with other immunosuppressants. Therefore the effect of CNI(r) with everolimus might be similar to the effect of CNI(s) when used with MPA.

   In order to account for this we calculated an additional estimate of the effect of the control, but this time compared to the putative placebo with standard dose CsA to the active control (i.e., Anti-IL2+CS+CsA(s) - Anti-IL2+CS+CsA(s)+MPA) using the same model as the sponsor. The estimate of Anti-IL2+CS+CsA(s) - Anti-IL2+CS+CsA(s)+MPA is 0.1715 (+/-0.0197), with a 95% CI of (0.1319, 0.2110) (results shown above in the last row of SAS output). The lower bound around the estimated difference of 13.2% is lower than the lower bound of 18.9% estimated when modeling the putative placebo with a reduced dose CsA; however, the results are consistent with those described above.

2) In the sponsor's literature search discussed above, there are 3 studies that assessed the effect of MPA in a regimen of CsA and CS. Additionally, there is a placebo controlled study summarized in the CellCept (MMF) label (these four studies are described below in Table 4). If the sponsor's assumption regarding additivity of

drug effects is true, then the estimate of the effect of MPA by comparing CS+CsA(s) - CS+CsA(s)+MPA should be an unbiased estimate of Anti-IL2+CS+CsA(s) - Anti-IL2+CS+CsA(s)+MPA. An estimate comparing CS+CsA(s)+AZA - CS+CsA(s)+MPA will lead to conservative results.

The benefit of this approach is that the estimate of the effect of MPA would be estimated using randomized studies that compared a regimen contained MPA to one not containing MPA, which is a more accurate way to estimate an effect.

Table 4: RCTs to Assess Effect of MMF/MPA to Regimen of CS+CsA(s)

Study	CS+CsA(s)+MMF	CS+CsA(s)+AZA	95% CI
Sadek	52/162 (32.1%)	73/157 (46.5%)	(-25.1,-3.8)
Study Group 1997	37/167 (22.2%)	67/166 (40.4%)	(-28.0, -8.4)
Study Group 1997	66/173 (38.2%)	83/166 (50.0%)	(-22.3, -1.3)
Pooled CI *			(-20.9, -9.1)
Study	**CS+CsA(s)+MMF**	**CS+CsA(s)**	**95% CI**
CellCept label*	50/165 (30.3%)	93/166 (56.0%)	(-15.4, -36.0)

* Dersimonian and Laird CI using inverse pooling method
** Numerators estimated
95% CI around risk difference

The one placebo study supports a margin of 10% since the lower bound of the CI is 15.4, which is greater than 10%. The pooled results versus AZA support a margin as large as 9.1. Note that this estimate is conservative since it estimates the effect of MPA over AZA and not the effect of MPA over placebo.

3) A final method is to estimate the effect of MPA using studies where MPA is used with tacrolimus rather than cyclosporine. Again the benefit with this approach is that it can measure the effect of MPA from within randomized studies. If the assumption is again used that the drug effects are additive, then the effect of MPA from studies with tacrolimus should be an unbiased estimate of the effect of MPA in a regimen with CsA and induction.

Table 5: RCTs to Assess Effect of MMF/MPA to Regimen of CS+Tacrolimus

Study	CS+tacro+MMF	CS+tacro	95% CI
Squifflet (6m)	4/71 (5.6%)	29/82 (35.4%)	(-41.4 -18.1)
Shapiro (12m)	27/102 (27%)	47/106 (44%)	(-30.6, -5.1)
Pooled CI*			(-35.7, -12.5)
Study	**CS+tacro+MMF**	**CS+tacro+AZA**	**95% CI**
Johnson (12m)	11/72 (15%)	13/76 (17%)	(-13.7, 10.0)
Busque (6m)	2/23 (9%)	8/23 (34.8%)	(-48.7, -3.5)
Mendez(6m) **	5/58 (8.6)	19/59 (32%)	(-37.5, -9.6)
Pooled CI*			(-32.5, 0.8)

* Dersimonian and Laird CI using inverse pooling method
**Results different than reference
95% CI around risk difference

As with method 2 above, the estimate of the effect of MMF is at least 12.5 and would support a margin of 10.0. The estimate of the effect of MMF over AZA does not support the margin, but is very conservative.

These three additional methods are all supportive of the proposed 10% margin.

Conclusion:

Given that there are no published studies that directly compared the active control (B+CS+CsA(s)+MPA) to the putative placebo (B+CS+CsA(r)), the sponsor's justification for the 10% margin for non-inferiority for study A2309 is based on a mixed effects modeling approach. While these models include several assumptions, this approach seems reasonable given the absence of comparative (concurrent) clinical data. The sponsor's main analysis resulted in a lower bound around the estimated difference in failure rates of 18.9% (e.g. the minimum (estimated) amount that active control is better than putative placebo in terms of efficacy failure). The sponsor's multiple sensitivity analyses yield more similar, yet larger except for one, estimates of the difference in treatment failure rates between active control and putative placebo and are therefore supportive of main comparison. Lastly, while the lower bound around the difference in failure rates was 18.9%, a non-inferiority margin greater than 10% (finding that the treatment is no worse than the active control by a maximum of 10%) would be unacceptable from a clinical point of view in renal transplantation.

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-21560	ORIG-1	NOVARTIS PHARMACEUTICALS CORP	CERTICAN (EVEROLIMUS) TABLETS

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

/s/

XIAO DING
12/10/2009

LAREE A TRACY
12/10/2009

KAREN M HIGGINS
12/10/2009

MOHAMMAD F HUQUE
12/11/2009

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-560/0000
Drug Name:	Everolimus Tablets, 0.25 mg, 0.5 mg, 0.75 mg, 1.0 mg (formerly SDZ, RAD, Certican)
Indication(s):	Prophylaxis of organ rejection in allogeneic kidney recipients
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submitted June 30, 2009
Review Priority:	Standard (6-month)
Biometrics Division:	Division of Biometrics VII
Statistical Reviewer:	John Stephen Yap, PhD
Concurring Reviewers:	LaRee Tracy, PhD
	Aloka Chakravarty, PhD
Medical Division:	Department of Special Pathogen and Transplant Products
Clinical Team:	Ergun Velidedeoglu, MD
	Joette Meyer, PharmD
Project Manager:	Jacquelyn Smith, MA
Keywords:	NDA review, clinical studies, noninferiority, safety

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

This statistical review and evaluation was performed in response to a consult from the Division of Special Pathogen and Transplant Products (DSPTP) for New Drug Application (NDA) 21-560/000 (received June 30, 2009) for everolimus tablets for the proposed indication of prophylaxis of organ rejection in allogeneic kidney recipients. This NDA submission contains 12-month results from study A2309, which was a Phase III, randomized, open-label, active-controlled, non-inferiority, 24-month study that evaluated the safety and efficacy of everolimus in *de novo* kidney transplantation. This statistical review will assess the following safety parameters measured in study A2309 (as requested in the DSPTP consult): glomerular filtration rate (GFR), proteinuria, and hyperlipidemia (total cholesterol, triglycerides, low density lipoprotein (LDL), high density lipoprotein (HDL) and cholesterol to HDL ratio. A separate statistical review of the efficacy of study A2309 was performed by Dr. Xiao Ding.

1.1 Conclusions and Recommendations

Study A2309 demonstrated that calculated 12-month GFR, using the modification of diet in renal disease (MDRD) formula, was similar between both everolimus groups and the Myfortic group (Table 1). Various sensitivity analyses, modeling and imputation methods for missing values resulted in similar results in 12-month GFR across treatment groups. Analyses of GFR trends found that the median GFR levels in the everolimus 1.5 mg group were numerically higher than those of Myfortic across most study visit windows but the treatment groups were not statistically significantly different at all time points.

Table 1: Calculated GFR* (MDRD) at 12 Months (ITT Population) by LOCF

	Everolimus 1.5 mg	Everolimus 3.0 mg	Myfortic
Sample Size	n=276	n=279	n=277
Mean (SD)	54.6 (21.7)	51.1 (22.8)	52.3 (26.5)
Median (Range)	55.0 (0.0-140.9)	51.5 (0.0-124.0)	50.1 (0.0-366.4)
Difference in Mean**	**2.4**	**-1.1**	
t-test based 95% CI	(-1.7,6.4)	(-5.2,3.0)	
t-test based 97.5% CI	(-2.3,7.0)	(-5.8,3.6)	
p-value (t-test)	0.2533	0.5938	
p-value (Wilcoxon)	0.0224	0.9895	

*GFR given in mL/min/1.73 m^2 ; ** Everolimus-Myfortic; LOCF=last-observation-carried-forward, graft loss imputed as zero (0); ITT=intent-to-treat

In both everolimus groups, there were statistically significant findings for proteinuria [as measured by urinary protein to urinary creatinine (UP/UC) ratio]. Specifically, the median UP/UC ratios for both everolimus arms were numerically higher than those of Myfortic at all measured time points after Day 3 (Days 7, 14 and Months 1 to 12). Differences in UP/UC ratio between the everolimus 1.5 mg and Myfortic groups were statistically significant at all time points from Month 1 until Month 12 (all p-values<0.05), except at Month 4 (p-value=0.11). Similarly, statistically significant differences between everolimus 3.0 mg and Myfortic were observed at Day 14 until Month 12 (all p-values<0.05). The differences in the median UP/UC

ratios between the everolimus 1.5 mg arm and Myfortic, which appeared to increase, beginning at 6-months post-transplant and continued through Month 12, appeared to be driven by differences between treatment groups in the subgroup of male study patients.

Total cholesterol levels were statistically significant and the medians were numerically higher in both everolimus arms compared to Myfortic from Month 1 post-transplant through Month 12 follow-up (p-values<0.05). From Month 1 onwards, the median total cholesterol values in both everolimus arms remained above 200 mg/dL, which is considered the clinical lower limit for hypercholesterolemia. Additionally, triglycerides levels were statistically significant and the medians were numerically higher in both everolimus arms compared to Myfortic from Month 1 through Month 12 follow-up (p-values<0.05). The median total cholesterol differences between the everolimus treatment arms and Myfortic appeared to be driven by differences between treatment groups in the subgroup of male study patients.

LDL and HDL levels were statistically significantly different between everolimus 1.5 mg and Myfortic at Month 1 (p-value=0.0086) and at Month 12 treatment endpoint (p-value=0.0158) for LDL and Month 6 (p-value=0.0013) and at Month 12 treatment endpoint (p-value 0.0002) for HDL. *Note: The Month 12 treatment endpoint was defined as the last post-baseline on-treatment observation up to and including the scheduled Month 12 visit.* Median post-baseline cholesterol to HDL ratios in the everolimus 1.5 mg arm were greater than in Myfortic except at Month 6, though treatment differences were not statistically significant. LDL levels were statistically significant at Month 1 (p-value=0.0080), Month 6 (p-value=0.0183), Month 12 (p-value=0.0022) and Month 12 treatment endpoint (p-value=0.0012) and the medians were numerically higher in the everolimus 3.0 mg group compared to Myfortic. Median HDL levels in the everolimus 3.0 group were numerically higher than in Myfortic but statistically significant differences between the two groups were only observed at the Month 12 treatment endpoint (p-value=0.0201).

This review provides detailed statistical analyses and assessment of the safety of everolimus with respect to the above described safety parameters. Refer to the clinical review, by Dr. Ergun Velidedeoglu, of NDA 21-560/000 details on the clinical assessment of these safety parameters.

1.2 Brief Overview of Clinical Studies

NDA 21-560/000 contains 12-month results for study A2309, which was a 24-month, Phase III, multi-center, randomized, open-label, non-inferiority study of everolimus for the indication of prophylaxis of organ rejection in *de novo* renal transplantation. The purpose of study A2309 was to evaluate the safety and efficacy of two therapeutic drug regimens of everolimus with reduced dose cyclosporine compared to a regimen of Myfortic with full dose cyclosporine. The study was conducted in 79 centers across Europe, North and South America. A total of 833 *de novo* kidney transplant male and female patients between the ages of 18 and 70 years were randomized in a 1:1:1 fashion to receive one of the following 3 treatment regimens:

- Everolimus 1.5 mg/day (0.75 mg bid) starting dose (target trough concentration of 3-8 ng/mL) with *reduced dose* Neoral + Simulect+corticosteroids (n=277)

- Everolimus 3.0 mg/day (1.5 mg bid) starting dose (target trough concentration of 6-12 ng/mL) with *reduced dose* Neoral + Simulect+corticosteroids (n=279)
- Myfortic 1.44 g (0.72 g bid) and standard dose Neoral + Simulect+corticosteroids (n=277)

The primary efficacy endpoint was efficacy failure at 12 months post transplantation, where efficacy failure is the composite endpoint consisting of treated biopsy-proven acute rejection (BPAR) episode (based on local laboratory assessment), graft loss, death, or loss to follow-up. The primary efficacy objective of study A2309 was to demonstrate that one or both everolimus regimens were non-inferior to the Myfortic regimen, based on a pre-specified 10% non-inferiority (NI) margin, in incidence of the primary efficacy failure endpoint. The main secondary efficacy endpoint was 12-month incidence of death, graft loss and loss to follow-up.

Note: Statistical analysis of the efficacy from study A2309 was reviewed separately by Dr. Xiao Ding.

The main safety endpoint was renal function at Month 12 as measured by the calculated GFR using the MDRD formula. The primary safety objective of the study was to demonstrate that at least one of the everolimus treatment arms was non-inferior to the Myfortic treatment arm by an 8 ml/min/1.73 m^2 pre-specified non-inferiority margin within 12 months of the initial dose of study medication with respect to the main safety endpoint.

Note: It is atypical to specify a NI margin for a safety variable particularly given the difficulty in justifying the NI margin. Therefore, the Agency considered results from the safety analysis along with efficacy in an overall risk benefit assessment.

The primary analysis for GFR was carried out in the intent-to-treat (ITT) population, defined as all randomized patients, using a t-test to compare the means at Month 12. The Hochberg procedure was utilized to adjust for multiple pair-wise comparisons. Several imputation approaches for missing 12-month GFR values were also performed as sensitivity analyses including imputation of zero (0) for graft loss and last-observation-carried-forward (LOCF) for death or loss to follow-up. Details of these analyses and results are provided in section 3.2.3.

1.3 Statistical Issues and Findings

The following issues were identified during the statistical safety review of study A2309:

- At Month 12, the incidence of premature treatment discontinuation in the everolimus 1.5 mg group, 3.0 mg and Myfortic groups was 30.0% (83/277), 34.1% (95/279), and 21.7% (60/277) respectively. Compared to the Myfortic group, the incidence was statistically significantly higher in the everolimus 1.5 mg group (p-value=0.03, Fisher's exact test) and in the everolimus 3.0 mg group (p-value=0.001, Fisher's exact test). Collection of adverse event data among patients who prematurely discontinued treatment was collected only up to 8 days following end of treatment. Due to the imbalance in rates of premature treatment discontinuation, with more discontinuation occurring in the everolimus groups, there may be

bias against the Myfortic regimen in rates of adverse events (Section 3.2.3 Patient Disposition, Demographic and Baseline Characteristics) (i.e. more adverse event reporting in the Myfortic group compared to the everolimus groups).

- The visit windows used in study A2309 to identify clinical measurements for particular time points were not uniform across the duration of the study and differed from those specified in the protocol. The visit windows given in Table 2 (Section 3.1.1 Study Design and Endpoints) and Table 3-1 of the Statistical Analysis Plan (SAP) are not consistent with what the protocol specified and do not have uniform widths. The protocol states that after Day 7, a visit window of 2 days up to Day 28, 1 week between Day 28 and Month 6, and 2 weeks after Month 6 is acceptable. Additionally, the start and end of each window is based on the midpoint between scheduled visits and some months did not have scheduled assessments. This resulted in some visit windows being longer than others. For example, the visit window for Month 2 was 30 days, for Month 6 was 45 days, and for Month 9 was 75 days. This may have led to some measurements being counted as part of a certain time point even when the measurements were not obtained near or during the specified time points. This may have resulted in less precision in measurements, particularly for those measurement obtained later on in the study follow-up. The overall impact this may have had on the final results are difficult to assess.

- The applicant used the last observation carried forward (LOCF) approach to impute missing observations in the 12-month analysis of GFR. The LOCF may have resulted in biased estimates for GFR because some observations were imputed at 12 months using values collected as early as Day 1. While this is a concern, results in 12-month GFR using LOCF generally did not differ significantly from the results using other imputation approaches (Section 3.2.2 Statistical Methods and Section 3.2.3.1 Glomerular Filtration Rate).

- The applicant's analysis that used LOCF to impute missing 12-month GFR was inconsistent. Specifically, for patients with reported Month 12 GFR values, these values were used in the analysis, regardless of whether the value was collected while the patient was on randomized treatment or off-randomized treatment. When Month 12 GFR values were missing, values from the last on-treatment GFR observations were used in the analysis. Therefore, the 12-month GFR analyses were comprised of both on-randomized treatment and off-randomized treatment GFR values. By definition, the preferred approach of LOCF should consist of the last observation collected, regardless of whether it is on-randomized treatment or not. This review repeated the 12-month GFR analysis following the definition of LOCF using the last-observation and the results were not significantly different from the applicant's results (Section 3.2.3.1 Glomerular Filtration Rate).

- The distributions in some clinical parameters (e.g. GFR, proteinuria, total cholesterol and triglycerides) at each time point were skewed or asymmetric due to extreme outlying observations and also possibly due to the amount of missing data, particularly at later time points. This data asymmetry was most pronounced in the urinary protein to urinary creatinine ratio measurements used to assess proteinuria. Use of statistical tests and models that assume a normal distribution of these asymmetric data are inappropriate. Therefore, nonparametric tests were utilized in this review. Additionally, sensitivity analyses were

performed that excluded extreme outlying values; however, the overall conclusions did not change (Section 3.2.3.2 Proteinuria and 5.1 Statistical Issues and Collective Evidence).

The following are the findings from the statistical safety review of study A2309:

- The mean (median) GFRs in the everolimus 1.5 mg arm were numerically higher than the mean (median) GFRs in both the everolimus 3.0 mg and Myfortic arms at 12 months post-transplant. The everolimus treatment arms compared to Myfortic were not statistically significantly different at some time points (Section 3.2.3.1 Glomerular Filtration Rate).

- The UP/UC ratio, total cholesterol and triglycerides levels were numerically higher in median values and statistically significantly different at most time points up to month 12 in both everolimus arms compared to the Myfortic arm (Section 3.2.3.2 Proteinuria and Section 3.2.3.3 Lipids).

- In subgroup analyses by gender, treatment differences were observed in the subgroup of male patients in the analyses of UP/UC ratios and total cholesterol. In male patients UP/UC ratio and total cholesterol were statistically significantly higher in both everolimus arms compared to Myfortic. The median total cholesterol levels of the male subgroup in the Myfortic arm were numerically lower than all other subgroups based on gender and treatment (Section 4.1 Proteinuria and Section 4.2 Lipids).

- LDL and HDL levels were statistically significantly different between everolimus 1.5 mg and Myfortic at Month 1 (p-value=0.0086) and at Month 12 TEP (p-value=0.0158) for LDL and Month 6 (p-value=0.0013) and at Month 12 TEP (p-value 0.0002) for HDL. Median post-baseline total cholesterol to HDL ratios in the everolimus 1.5 mg arm were greater than in Myfortic except at Month 6, though treatment differences were not statistically significant. Post-baseline LDL and HDL levels were statistically significantly different between everolimus 3.0 mg and Myfortic at Month 1 (p-value=0.0080), Month 6 (p-value=0.0183), Month 12 (p-value=0.0022) and at Month 12 TEP (p-value=0.0012) for LDL and Month 12 TEP (p-value=0.0201) for HDL. Median post-baseline total cholesterol to HDL ratios in the everolimus 3.0 mg arm were greater than in Myfortic and treatment differences were statistically significant at Month 12 (p-value=0.0013) and Month 12 TEP (p-value=0.0142) (Section 3.2.3.3 Lipids and Section 4.2 Lipids).

- Statistically significantly more patients in the everolimus arms prematurely discontinued treatment at month 12 compared to the rate observed in the Myfortic arm. This may have led to discordant rates of reported adverse events given that adverse events were systematically collected only up to 8 days following treatment discontinuation (Section 3.2.3 Patient Disposition, Demographic and Baseline Characteristics).

- There was an increasing number of missing data as study follow-up time increased. For UP/UC ratios and lipid levels, starting at Day 14, about 90% of data were collected compared to only about 70% at Month 12 (Section 3.2.3.2 Proteinuria and Section 3.2.3.3 Lipids).

2. INTRODUCTION

2.1 Overview

The applicant (Novartis Pharmaceuticals) previously submitted the original NDA 21-560 on December 19, 2002, which included two 24- month pivotal, Phase III, multi-center, randomized, double-blind, double-dummy, and parallel group studies (studies B201 and B251) of fixed dose everolimus and standard dose Neoral (cyclosporine A, CsA). The original NDA was issued an Approvable action (former terminology for a Complete Response) on October 20, 2003 citing an insufficiently safe regimen for everolimus when used with full-dose cyclosporine despite demonstration of efficacy. The NDA Action letter required both an effective and safe dosing regimen of everolimus and cyclosporine for the prophylaxis of organ rejection in renal transplant patients for approval. In response to the NDA Approval letter, the applicant submitted two supplementary studies (studies A2306 and A2307) on February 27, 2004 to NDA 21-560. Studies A2306 and A2307 evaluated therapeutic drug monitoring (TDM) of everolimus and reduced dose Neoral; however these studies were limited in that they did not include a non-everolimus control regimen. A second Approval letter was issued on August 27, 2004 citing that the results from the supplementary studies were not acceptable to support the approval of the NDA due to study design limitations.

This current review covers the applicant's resubmission on June 30, 2009 to NDA 21-560/000 in support of the efficacy and safety of everolimus for the indication of prophylaxis of organ rejection in *de novo* renal transplantation. The applicant submitted the 12-month results from study A2309, a Phase III, 24-month, multi-center, randomized, open-label, non-inferiority study comparing two everolimus doses (1.5 mg and 3.0 mg starting dose) with TDM and reduced dose Neoral versus Myfortic and standard dose Neoral.

This review covers a detailed safety review of study A2309 in response to a safety consult from DSPTP on July 16, 2009. The consult requested a statistical evaluation of (1) the trends of GFR, proteinuria, and hyperlipidemia levels (2) population subgroups that may affect the levels in (1) and (3) potential association between GFR and proteinuria levels. Study A2309 evaluated two everolimus regimens; however, the applicant is only seeking approval for everolimus 1.5 mg/day (target trough concentration of 3-8 ng/mL) with reduced dose cyclosporine regimen. Therefore, this review will focus on the comparison of everolimus 1.5 mg and control Myfortic; however there are references or results of comparison between the everolimus 3.0 mg/day regimen and Myfortic when necessary for completeness. The investigational drug will be referred to here using the generic name everolimus because a final trade (innovator) name has not been approved at the time of the review.

2.2 Data Sources

Study A2309 data sets analyzed in this review, including responses from the applicant to FDA requests, are located in the CDER Electronic Document Room (EDR) at the following link:

\\Cdsesub1\evsprod\NDA021560\0010\m5\datasets\rad001a2309\analysis
\\Cdsesub1\evsprod\NDA021560\0016

\\Cdsesub1\evsprod\NDA021560\0036

The study A2309 clinical study report, including the protocol and SAP, is located in the EDR at the following link:

\\Cdsesub1\evsprod\NDA021560\0010\m5\53-clin-stud-rep\535-rep-effic-safety-stud\prophylaxis-of-kidney-transplant-rejection\5351-stud-rep-contr\rad001a2309

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

This review focuses on specific safety parameters measured in study A2309. For a complete statistical evaluation of the efficacy results from study A2309, please refer to the review authored by Dr. Xiao Ding.

3.2.3 Study Design and Endpoints

Study A2309 was a Phase III, 24-month, multicenter, multi-national, randomized, open-label, non-inferiority study in *de novo* renal transplant patients. Patients were randomized to receive either one of two regimens of concentration-controlled everolimus (1.5 and 3.0 mg/day initial doses) with titrated reduced dose Neoral or 1.44 g Myfortic with titrated standard dose Neoral in a 1:1:1 fashion, as shown in Figure 1. Patients in each study arm also received Simulect (basiliximab) (for antibody induction therapy) and corticosteroids. Everolimus doses were adjusted to reach blood trough level targets of 3-8 ng/mL (for everolimus 1.5 mg) and 6-12 ng/mL (for everolimus 3.0 mg). Neoral doses were adjusted to get trough blood level values within the pre-specified target ranges as shown in Table 2.

Table 2: Protocol-Specified Neoral (cyclosporine) Trough Values (mg/mL)

Groups	Month 1	Starting Month 2	Starting Month 4	Starting Month 6
Everolimus	100-200	75-150	50-100	25-50
Myfortic	200-300	-	150-250	-

Eligible patients included male and female renal transplant patients, age 18 to 70 years, who received a primary cadaveric kidney from living unrelated or non-HLA identical living related donor.

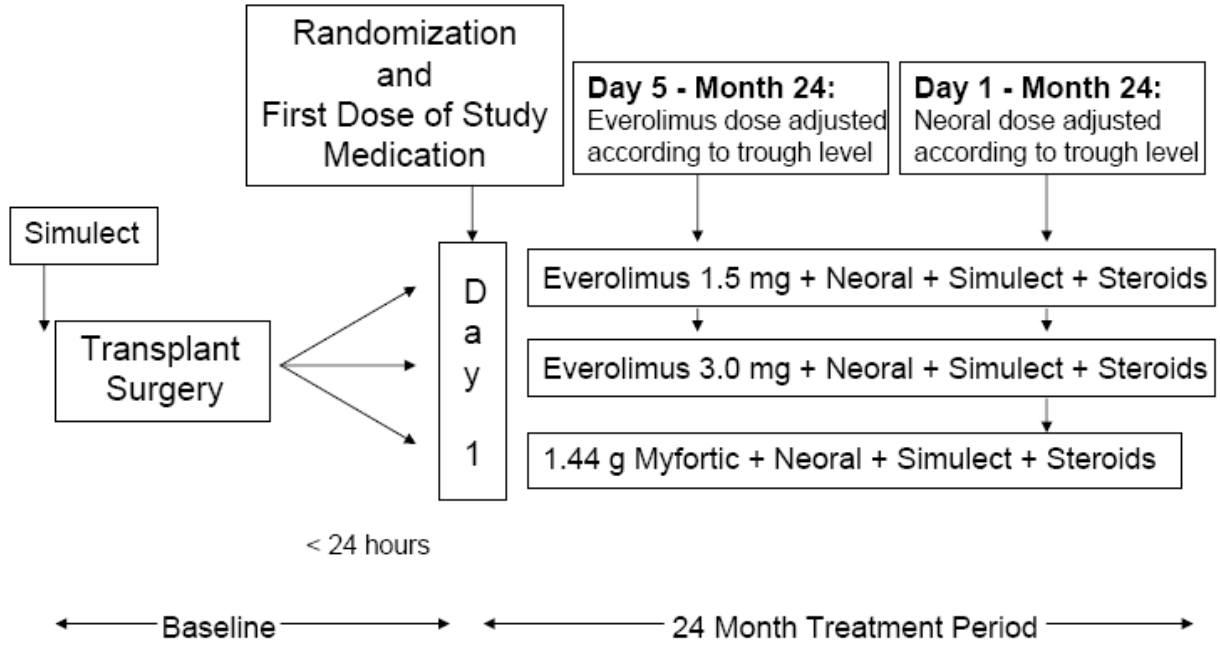
Study populations as defined in the protocol include:

The i**ntent-to-treat (ITT) population** consists of all patients randomized after transplantation.

The **safety population** consists of all patients that received at least one dose of study drug and had at least one post-baseline safety assessment.

The **per-protocol (PP) population** consists of all randomized patients who took study treatment according to the protocol without any major deviations from the protocol procedures.

Figure 1: Study Design*



*Based on Figure 4-1 of the applicant's Clinical Overview.

As stated in the protocol, efficacy analysis was to be performed in the ITT and PP populations whereas safety analysis was to be performed in the safety population. The primary efficacy population was the ITT.

Reviewer's Comments: *The applicant also performed analysis on the* ***safety on-treatment population*** *although the protocol did not explicitly define this population. However, the SAP defined an on-treatment observation as any assessment obtained on and after Day 1 but no later than two days after the discontinuation of randomized study medication. Therefore, the definition of safety on-treatment population is taken to be the definition of safety population and on-treatment observation combined.*

Baseline assessments occurred in the time period starting 24 hours prior to transplantation until the time of randomization. Patients were assessed at Baseline, Days 1, 3, 4, 5, 7, 14, 28 and Months 2, 3, 4, 6, 7, 9, 12, 18, and 24. The time windows of visits for analysis purposes are shown in Table 3, as given in the SAP. The date of first dose of study medication was Day 1. The cutoff date for the 12-month analysis was Day 450.

The applicant considered the following summary of a patient's last observation post-baseline:

Month 12 Treatment Endpoint (TEP) = Last post-baseline *on-treatment* observation up to and including the scheduled Month 12 visit.

Month 12 Study Endpoint (SEP) = Last post-baseline observation up to and including the scheduled Month 12 visit. This includes assessments obtained after discontinuation of treatment.

Table 3: Re-aligned Visit Windows*

Visit Number	Visit Name	Starting Day of Window	Ending Day of Window
0	Pre-baseline	Day -60	Day -8
1	Baseline**	Day -7	Up to the first dose of study medication
2	Day 1	Day 1***	Day 2
3	Day 3	Day 3	Day 4
4	Day 7	Day 5	Day 11
5	Day 14	Day 12	Day 21
6	Month 1	Day 22	Day 44
7	Month 2	Day 45	Day 75
8	Month 3	Day 76	Day 105
9	Month 4	Day 106	Day 150
10	Month 6	Day 151	Day 195
11	Month 7	Day 196	Day 240
12	Month 9	Day 241	Day 315
13	**Month 12**	**Day 316**	**Day 450**
32	**Month 12 TEP**	**Day 1**	**Day 450**
42	**Month 12 SEP**	**Day 1**	**Day 450**
14	Month 18	Day 451	Day 630
15	Month 24	Day 631	Day 810

*Based on Table 3-1 of SAP; **prior to first dose of study medication; *** the first dose of study medication; TEP: treatment endpoint; SEP: study endpoint; rows in bold letters were used in the analysis of renal and lab data.

Reviewer's Comment: *The visit windows given in Table 3 are inconsistent with what the protocol specified and do not have uniform widths. The protocol states that after Day 7, a visit window of 2 days up to Day 28, 1 week between Day 28 and Month 6, and 2 weeks after Month 6 is acceptable. Additionally, because the start and end of each window is based on the midpoint between scheduled visits and some months did not have scheduled assessments, some windows were longer than others. For example, the visit window for Month 2 was 30 days, for Month 6 was 45 days, and for Month 9 was 75 days. This will result in some measurements being counted as being part of a certain time point even when they are far off from that time point.*

Central laboratory values for serum creatinine were used for all renal function analysis. Local laboratory serum creatinine values were used when the central laboratory values were missing.

Patients who prematurely discontinued study medication prior to the 24-month treatment period were to be contacted at Months 3, 6, 9, 12, 18 and 24 visits to obtain follow-up information. Among patients who prematurely discontinued treatment, adverse events/infections were collected only up to Day 7 following end of treatment. During the review, the applicant submitted an updated analysis (received on October 5, 2009) including all adverse

events/infections with onset date up to 30 days after treatment discontinuation. Serious adverse events were collected up to 30 days following treatment discontinuation.

The primary efficacy endpoint was a composite consisting of treated **BPAR episode (based on local assessment), graft loss, death or loss to follow-up** at 12 months post transplant in the ITT population.

The primary objective of study A2309 was to demonstrate that at least one of the everolimus treatment arms was not worse than (non-inferior to) the Myfortic treatment arm within 12 months of the initial dose of study medication with respect to the primary composite efficacy endpoint. The pre-specified non-inferiority efficacy margin was 10%.

The main secondary efficacy (composite) objective was to compare the incidence of graft loss, death or loss to follow-up in the ITT population between the everolimus and Myfortic treatment arms at 12-months post-transplantation.

3.2.3 Statistical Methods

Please refer to the statistical review authored by Dr. Xiao Ding for details on the efficacy analyses.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

A total of 833 patients (277 in everolimus 1.5 mg arm, 279 in everolimus 3.0 mg arm, and 277 in Myfortic arm) were randomized into the study, as shown in Table 4. One patient in the everolimus 1.5 mg arm withdrew consent after randomization but was still counted in the ITT population. There were eight patients (3 randomized to everolimus 1.5 mg, 1 randomized to everolimus 3.0 mg, and 4 randomized to Myfortic) who never received study drug (7 patients did not receive 2 doses of Simulect and 1 patient's course of study drug was less than 6 months). The PP population contained approximately 73% to 83% of all randomized patients in all treatment groups.

Table 4: A2309 Study Populations

	Number of Patients (%)*			
Population	**Everolimus 1.5 mg**	**Everolimus 3.0 mg**	**Myfortic**	**Total**
Intent-to-Treat	277	279	277	833
Safety	274 (99)	278 (99)	273 (99)	825 (99)
Per-Protocol	215 (78)	205 (73)	230 (83)	650 (78)

*% relative to the Intent-to-Treat

Patients ranged from 18-70 years in age (mean=46 years) with about two-thirds of patients being male in all treatment groups as shown in Table 5. Approximately two-thirds of patients were Caucasian and only about 14% were Black; the rest were either Asian (13%) or classified as Other race (5%). Twenty one percent of patients were diabetic at baseline. Fifty-two percent of organ donors were male and 68% were less than 50 years of age. There were no differences among treatment groups with respect to baseline characteristics and demographics.

The study was performed across 79 centers located across Europe, North America and South America. The average enrollment was 10 patients per center with a range of 1 to 95. The countries (number of centers) in the study were: Argentina (3), Australia (8), Brazil (4), Canada (1), Hong Kong (1), Italy (4), New Zealand (1), Singapore (1), Slovakia (2), South Africa (1), South Korea (6), Sweden (1), Taiwan (1), Turkey (3), United Kingdom (3), and United States (39).

Table 5: Demographic and Baseline Characteristics

	Everolimus 1.5 mg n=277	Everolimus 3.0 mg n=279	Myfortic n=277	Total n=833
Age at Baseline (yrs)				
Mean	46	45	47	46
Range	18-70	18-70	18-70	18-70
Sex				
Male	176 (63.5)	191 (68.5)	189 (68.2)	556 (66.7)
Female	100 (36.1)	88 (31.5)	88 (31.8)	276 (33.1)
Missing	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.2)
Race				
Caucasian	193 (69.7)	180 (64.5)	190 (68.6)	563 (67.6)
Black	34 (12.3)	40 (14.3)	39 (14.1)	113 (13.6)
Asian	32 (11.6)	38 (13.6)	36 (13.0)	106 (12.7)
Native American	0 (0.0)	5 (1.8)	1 (0.4)	6 (0.7)
Pacific Islander	0 (0.0)	1 (0.4)	0 (0.0)	1 (0.1)
Other	17 (6.1)	15 (5.3)	11 (4.0)	43 (5.2)
Diabetic	58 (20.9)	48 (17.2)	69 (24.9)	175 (21.0)
Sex of Donor				
Male	154 (55.6)	139 (49.8)	136 (49.1)	429 (51.5)
Female	122 (44.0)	140 (50.2)	140 (50.5)	402 (48.3)
Missing	1 (0.4)	0 (0.0)	1 (0.4)	2 (0.2)
Age of donor				
<50	181 (65.3)	203 (72.8)	182 (65.7)	566 (67.9)
>=50	95 (34.3)	76 (27.2)	94 (33.9)	265 (31.8)
Missing	1 (0.4)	0 (0.0)	1 (0.4)	2 (0.2)

There were significantly more patients in the everolimus 1.5 mg arm (30%) and everolimus 3.0 mg (34.1%) compared to Myfortic (22%), (everolimus 1.5 mg v. Myfortic, p=0.03; everolimus 3.0 mg v. Myfortic, p=0.001 Fisher's exact) who prematurely discontinued treatment (Table 6). The main reasons for this was adverse event related (18% for everolimus 1.5 mg versus 9.4% Myfortic) and withdrawal of consent (4% for everolimus 1.5 mg versus 2% Myfortic). Study discontinuation due to withdrawal of consent was also higher in the everolimus 1.5 mg arm compared to Myfortic (7% versus 4%, respectively). For more details about treatment and study discontinuation, please refer to the statistical review authored by Dr. Xiao Ding.

Table 6: Patient Disposition

	Everolimus 1.5 mg n=277 (%)	Everolimus 3.0 mg n=279 (%)	Myfortic n=277 (%)	Total n=833 (%)
Premature treatment discontinuation	83 (30.0)#	95 (34.1)^	60 (21.7)	238 (28.6)
Adverse Event	50 (18.1)	57 (20.4)	26 (9.4)	133 (16.0)
Unsatisfactory therapeutic effect	11 (4.0)	14 (5.0)	13 (4.7)	38 (4.6)
Withdrew Consent	11 (4.0)	4 (1.4)	5 (1.8)	20 (2.4)
Graft Loss	3 (1.1)	6 (2.2)	6 (2.2)	15 (1.8)
Death	3 (1.1)	3 (1.1)	4 (1.4)	10 (1.2)
Other*	5 (1.8)	11 (3.9)	5 (2.2)	21 (2.5)
Study discontinuation	36 (13.0)	27 (9.7)	25 (9.0)	88 (10.6)
Withdrew Consent	20 (7.2)	8 (2.9)	12 (4.3)	40 (4.8)
Death	7 (2.5)	9 (3.2)	6 (2.2)	22 (2.6)
Graft Loss	9 (3.2)	10 (3.6)	7 (2.5)	26 (9.4)

*administrative problems, abnormal laboratory values, abnormal tests, protocol violation
#everolimus 1.5 mg v. Myfortic, p-value=0.030
^everolimus 3.0 mg v. Myfortic, p-value=0.001

3.2 Evaluation of Safety

This safety review of study A2309 consists of a focused evaluation of GFR, proteinuria, lipid elevation, and the potential association between GFR and proteinuria, including subgroup and outlier analyses. All analyses are between randomized treatment regimens.

3.2.3 Endpoints

The main safety endpoint in study A2309 was **renal function** as measured by the calculated GFR using the MDRD formula.

The main safety objective was to demonstrate that similar renal function (as measured by the calculated GFR using the MDRD formula) was achieved in the everolimus treatment arms compared to the Myfortic treatment arm at 12 months post-transplantation. The null and alternative hypotheses were as follows:

Null hypothesis: the mean GFR of the everolimus arm is lower (worse) than that of the Myfortic arm by 8 mL/min/1.73 m^2 or more.

Alternative hypothesis: the mean GFR of the everolimus arm is not lower (not worse) than that of the Myfortic arm by 8 mL/min/1.73 m^2 or more.

Renal function was also assessed by measuring proteinuria levels as determined by a spot UP/UC ratio.

In study A2309, the assessment of safety was based mainly on the frequency of adverse events (including infections and serious adverse events) and on laboratory values that were beyond pre-specified ranges. Safety clinical parameters assessed in this review included total cholesterol, triglycerides, LDL, HDL and HDL to cholesterol ratio

3.2.3 Statistical Methods

Study A2309 compared calculated GFR in two everolimus arms (1.5 mg and 3.0 mg starting dose) to calculated GFR in the Myfortic arm in the ITT population. To control for multiple comparisons, the applicant used the modified Bonferoni testing procedure proposed by Hochberg (1988)[1] to maintain the overall Type I error rate at α=0.05: T-test based, two-sided 95% and 97.5% confidence intervals (CI) were computed for the mean calculated GFR at 12 months post-transplantation between the everolimus and Myfortic arms. If both comparisons yielded two-sided 95% CIs whose lower confidence limits were less than -8 mL/min/1.73 m^2 then both everolimus arms were claimed to be not worse than (non-inferior to) the Myfortic arm. Otherwise, the everolimus arm whose comparison with the Myfortic arm yielded a 97.5% CI with a lower limit above -8 mL/min/1.73 m^2 was claimed to be not worse than (non-inferior to) the Myfortic arm.

Reviewer's Comment: The non-inferiority margin of 8 mL/min/1.73 m^2 *is not considered valid as its determination was not justified using historical data. Safety was therefore to be evaluated along with efficacy in an overall risk/benefit assessment.*

Methods for Imputing Missing Values for GFR

The treatment arms comparison was made on the ITT population with the following imputation methods for patients with missing 12-month values:

1. For graft-loss, impute zero (0).
2. For death or loss to follow-up, impute using LOCF.

Reviewer's Comment: *Although the LOCF imputation approach was pre-specified in the protocol, this approach can lead to biased point estimates and variance. This is especially problematic when the values imputed for Month 12 were from earlier months e.g. Months 1 to 5.*

The applicant considered two endpoints at Month 12 using the LOCF method:

1. End of Treatment (up to Month 12): the last post-baseline *on-treatment* observation of GFR up to and including the scheduled Month 12 visit.
2. End of Study (up to Month 12): the last post-baseline observation of GFR, including values observed during follow-up visits after discontinuation of study medication, during the 12-month study period.

[1] Hochberg, Y. (1988). A Sharper Bonferroni procedure for Multiple Tests of Significance. *Biometrika* 75: 383-386.

Table 7 was based on Table 8-2 of the SAP and summarizes all the imputation methods used by the applicant.

Table 7: Imputation Methods for Missing Month 12 GFR*

Method #	Description of 12-Month GFR Missing Value Imputation Method
Method 1	graft-loss = 0; death or lost to follow up for renal function = LOCF1 (last-observation-carried-forward approach 1: End of Treatment (up to Month 12))
Method 2	graft-loss = 0; death or lost to follow up for renal function = LOCF2 (last-observation-carried-forward approach 2: End of Study (up to Month 12))
Method 3	exclude all patients with missing 12-month value due to graft loss, death or loss to follow up for renal function
Method 4	graft-loss = 0; death or lost to follow up for renal function at or after 6-month visit = LOCF 1 (last-observation-carried-forward approach 1: End of Treatment (up to Month 12)); exclude all patients with missing 12-month value due to death or loss to follow up for renal function prior to Month 6
Method 5	all drop-out patients (graft-loss, death or lost to follow up for renal function) = LOCF1 (last-observation-carried-forward approach 1: End of Treatment (up to Month 12))

*Based on Table 8-2 of SAP.

Reviewer's Comment: The applicant considered Method 1 in Table 7 as the primary analysis for GFR. However, the reviewer considered Method 2 to be the primary one because the imputation by LOCF uses the last observation, regardless of whether it is on-treatment or not.

The applicant calculated summary statistics (mean, standard deviation, median, minimum and maximum) and mean plots for GFR, UP/UC ratio, and lipids measurements over time. Marginal distributions at each visit window were compared using the nonparametric Wilcoxon rank-sum test.

For GFR, the applicant also performed the following analyses:

- Sensitivity analyses using other imputation methods
- Comparison of mean using analysis of covariance (ANCOVA) adjusting for covariates
- Exploratory analysis using mixed effects model
- Categorical analysis at Months 1 and 12, including shift tables (whereby categorical GFR values at Month 1 were compared with categorical GFR values at Month 12), using the National Kidney Foundation (NKF) criteria

This review repeated the categorical analysis that was done by the applicant but using smaller categories to determine whether the data are uniformly distributed within the larger categories or not. The categories used by the applicant were <30, 30 to <60 and ≥60 whereas the categories used here were <30, ≥ 30 to < 50, 50 to <55, 55 to <60 and ≥60.

In the analysis of UP/UC ratios, the applicant performed mixed effects model and categorical analyses using the National Kidney Foundation (NKF) criteria defined as follows: Normal (<30

mg/g; <3.39 mg/mmol), Mild (30-<300 mg/g; 3.39-<33.9 mg/mmol), Sub-nephrotic (300-<3000 mg/g; 33.9-<339 mg/mmol) and Nephrotic proteinuria (≥3000 mg/g; ≥339 mg/mmol).

This review includes an analysis that replicates the applicant's categorical analysis of UP/UC ratios using the following clinically relevant categories (in unit of grams/grams) that were suggested by the clinical review team: Normal (≤0.2), Mild (>0.2 to <1), Sub-nephrotic (1 to <3) and Nephrotic (≥3). Analysis was also performed using shift tables similar to the analysis for GFR.

The analyses of total cholesterol by the reviewer consisted of categorical analysis (<200 mg/dL, 200 to <240 mg/dL and ≥240 mg/dL). These categories are consistent with the American Heart Association guidelines for detection of high cholesterol.

To account for potential non-normal or asymmetric marginal distributions of the data at each visit window, median plots for all measurements were evaluated by the reviewer.

3.2.3 Results and Conclusions

3.2.3.1 Glomerular Filtration Rate

At 12-months post-transplantation, the mean (SD) GFR (mL/min/1.73 m^2) in the everolimus 1.5 mg, everolimus 3.0 mg and Myfortic groups were 54.6 (21.7), 51.1 (22.8) and 52.3 (26.5), respectively as shown in Table 8.

Using the protocol defined last observation carried forward (LOCF) imputation approach (primary imputation approach) and with no imputation, mean GFR values were similar across treatment groups. In the LOCF analysis, patients with a graft loss were considered as having a GFR of zero (0), while those who either died or were lost to follow-up had their last value used. Additional methods (as described in section 3.2.2 Statistical Methods) for imputation were also used and similar results were obtained. P-values from the Wilcoxon rank-sum test are also reported and show that the GFR values were statistically significantly different between everolimus 1.5 mg and Myfortic but not between everolimus 3.0 mg and Myfortic (Applicant's Analysis and LOCF).

Reviewer's Comment: *The applicant's analysis in Table 8 was replicated from Table 5-1 of the applicant's Clinical Overview. The applicant used the variable "gfr_m1", defined as imputation method 1 treatment endpoint, in the "renal.xpt" dataset. The Agency requested clarification of this variable in a letter to the applicant dated November 13, 2009. The applicant clarified that "gfr_m1" was based on the Month 12 GFR value imputed from the last* ***on-treatment*** *observation if the Month 12 GFR was missing or the Month 12 value if it was available, regardless of whether the Month 12 value was an on-treatment value or not. The reviewer's preference is that the LOCF approach should impute values using the last observation, regardless of whether it is on-treatment or not, in order to preserve the consistency of the analysis when combining with available 12-month values. This analysis is provided above in Table 8 for the LOCF analysis.*

Table 8: Calculated GFR* (MDRD) at 12 Months (ITT Population)

	Everolimus 1.5 mg	Everolimus 3.0 mg	Myfortic
ITT Population Size	n=277	n=279	n=277
Applicant's Analysis**	n=275	n=278	n=277
Mean (SD)	54.6 (21.7)	51.3 (22.7)	52.2 (26.7)
Median (Range)	55.0 (0.0-140.9)	51.6 (0.0-124.0)	49.7 (0.0-366.4)
Difference in Means***	**2.4**	**-0.9**	
t-test based 95% CI	(-1.7,6.4)	(-5.0,3.2)	
t-test based 97.5% CI	(-2.3,7.0)	(-5.6,3.8)	
p-value (t-test)	0.2513	0.6727	
p-value (Wilcoxon)	0.0222	0.9232	
LOCF****	n=276	n=279	n=277
Mean (SD)	54.6 (21.7)	51.1 (22.8)	52.3 (26.5)
Median (Range)	55.0 (0.0-140.9)	51.5 (0.0-124.0)	50.1 (0.0-366.4)
Difference in Mean***	**2.4**	**-1.1**	
t-test based 95% CI	(-1.7,6.4)	(-5.2,3.0)	
t-test based 97.5% CI	(-2.3,7.0)	(-5.8,3.6)	
p-value (t-test)	0.2533	0.5938	
p-value (Wilcoxon)	0.0224	0.9895	
No Imputation	n=245	n=244	n=248
Mean (SD)	56.3 (20.1)	55.0 (19.8)	54.4 (26.4)
Median (Range)	55.3 (4.6-140.9)	53.8 (8.7-124.0)	50.8 (6.8-366.4)
Difference in Mean***	**2.0**	**0.5**	
t-test based 95% CI	(-2.2,6.1)	(-3.5,4.7)	
t-test based 97.5% CI	(-2.8,6.7)	(-4.1,5.3)	
p-value (t-test)	0.3602	0.7810	
p-value (Wilcoxon)	0.0573	0.4795	

*GFR given in mL/min/1.73 m^2 ; **See Reviewer's Comment; *** Everolimus-Myfortic; **** LOCF = last-observation-carried-forward, graft loss imputed as zero (0)

Reviewer's Comment: *The LOCF approach may have resulted in biased estimates for GFR because some observations were imputed at 12 months using values collected as early as Day 1. The frequencies of Month 12 imputations from earlier visit windows are presented in Table 9. Table 9 shows that for last on-treatment observations, there were more imputations at Month 12 for both everolimus groups compared to Myfortic from earlier visit windows. For example, there were 10 values at Day 7 that were imputed at Month 12 for the everolimus 1.5 mg compared to only 7 for Myfortic. Last observations that were not necessarily on-treatment were more uniformly distributed across visit windows for all treatment groups.*

The applicant's analysis using ANCOVA and mixed effects models resulted in similar conclusions. For ANCOVA, the fixed effects in the final model included the Month 1 GFR (defined as baseline), age at baseline, donor's age, graft type (living or cadaveric) and body mass index (BMI). For the mixed effects model, the covariates in the final model include the number of days since the start of the first dose, cyclosporine trough level, treatment group, age of donor,

patient gender, diabetes status, graft status (living or cadaveric), delayed graft function indicator and BMI.

Table 10 shows the summary statistics for GFR across selected visit windows in the ITT population for everolimus 1.5 mg and Myfortic. These data are also plotted in Figures 2 and 3 illustrating that the means and medians for the everolimus 1.5 mg group were consistently higher than those of Myfortic throughout the study including the Month 12 TEP and SEP, shown as Months 13 and 14, respectively, in both figures.

Table 9: Frequencies of Month 12 Imputations from Earlier Visit Windows

	Last On-treatment Observation				Last Observation			
Visit Window	EVR 1.5 mg	EVR 3.0 mg	Myfortic	Total	EVR 1.5 mg	EVR 3.0 mg	Myfortic	Total
Day 1	2	1	2	5	2	1	0	3
Day 3	1	3	3	7	1	1	3	5
Day 7	10	12	7	29	2	4	1	7
Day 14	9	6	4	19	3	3	3	9
Month 1	12	14	9	35	3	3	2	8
Month 2	11	12	6	29	4	0	2	6
Month 3	8	12	4	24	5	2	0	7
Month 4	8	11	4	23	3	1	1	5
Month 6	5	7	3	15	2	5	3	10
Month 7	5	7	10	22	1	6	3	10
Month 9	9	12	11	32	5	9	11	25
Month 12	194	181	209	584	245	244	248	737
Total	274	278	272	824	276	279	277	832

EVR=Everolimus

The everolimus 1.5 mg and Myfortic treatment groups were statistically significantly different (based on the Wilcoxon rank-sum test) at Months 1 (p-value=0.0371), 6 (p-value=0.0135), 7 (p-value=0.0153), 9 (p-value 0.0228), 12 TEP (p-value 0.0412) and 12 SEP (p-value 0.0324). There were no statistically significant differences between the everolimus 3.0 mg group and the Myfortic group at any visit windows. *Note: These multiple comparisons are unadjusted.*

Calculated GFR, by clinically relevant categories, at Months 1 and 12 in the everolimus 1.5 mg and Myfortic groups are presented in Table 11. The following observations can be drawn from results presented in Table 11:

- At Months 1 and 12, the marginal categories with the largest proportions of patients were GFR ≥60 and GFR ≥30 to <50 (mL/min/1.73 m^2) in the everolimus 1.5 mg and Myfortic groups.
- At Month 1, there were more patients with a calculated GFR falling into a high category in the everolimus 1.5 mg arm compared to the Myfortic arm (e.g. in the GFR ≥60 category: 45% everolimus 1.5 mg vs. 37% Myfortic). Among patients with a GFR ≥60, 63% and 60% in the everolimus 1.5 mg and Myfortic groups respectively retained that level at 12 months.

- At Month 1, only a few patients (7-8%) in the <30 category for both everolimus 1.5 mg and Myfortic groups remained at that level at Month 12 while the rest improved to higher categories.
- At Month 12, there were many more patients in the everolimus arms who had GFR ≥60 (39%) compared to Myfortic (30%) and there were less patients in the everolimus arms (29%) who had GFR in the ≥30 to <50 category compared to Myfortic (40%).

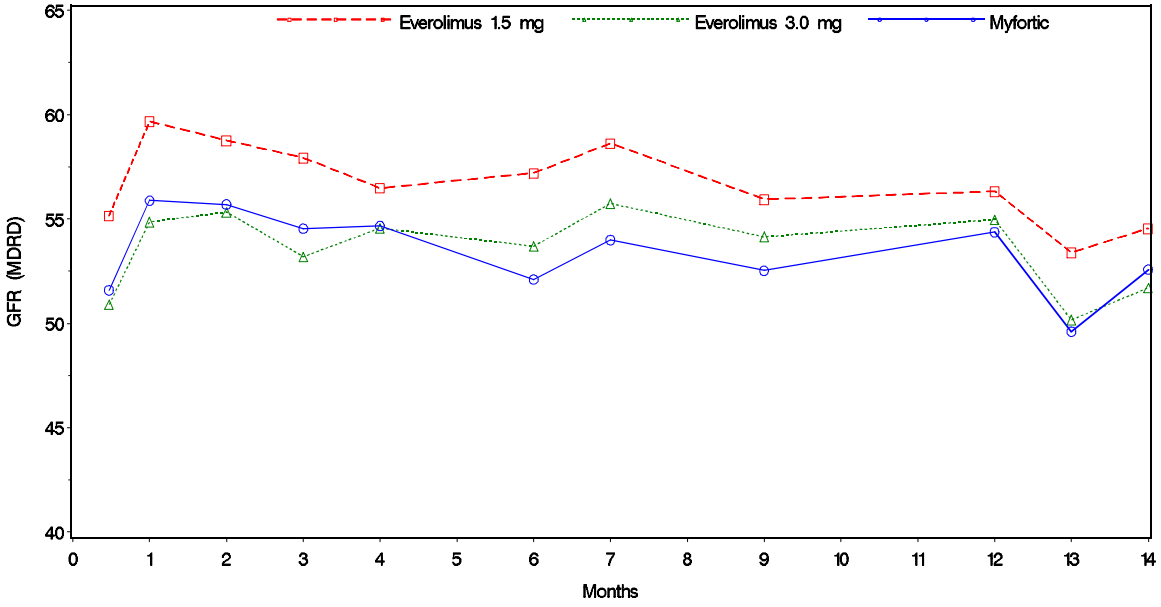
Note that Table 10 only includes data from patients who had both Months 1 and 12 GFR measurements, which was approximately 84% of the ITT population for both treatment arms.

Table 10: Calculated GFR* (MDRD) by Visit Window (ITT Population)

Visit Window	Treatment Group	n (%)***	Mean (SD)	Median (Range)	p-values** vs Myfortic
Day 14	EVR 1.5 mg	257 (93)	55.1 (25.4)	54.0 (3.5-199.3)	0.0561
	EVR 3.0 mg	252 (90)	50.9 (24.4)	50.8 (4.3-124.9)	0.9362
	Myfortic 1.44 g	251 (91)	51.6 (23.5)	49.7 (4.1-151.3)	
Month 1	EVR 1.5 mg	251 (91)	59.7 (22.5)	56.9 (6.8-181.2)	0.0291
	EVR 3.0 mg	252 (90)	54.9 (21.4)	54.6 (6.5-137.1)	0.7128
	Myfortic 1.44 g	251 (91)	55.7 (21.2)	53.0 (5.1-153.9)	
Month 3	EVR 1.5 mg	253 (91)	57.9 (20.8)	54.7 (8.9-175.9)	0.0768
	EVR 3.0 mg	252 (90)	53.2 (19.7)	51.9 (7.1-119.0)	0.4296
	Myfortic 1.44 g	256 (92)	54.5 (20.2)	52.2 (5.9-133.2)	
Month 6	EVR 1.5 mg	232 (84)	57.2 (19.9)	53.6 (6.4-138.4)	0.0129
	EVR 3.0 mg	241 (86)	53.7 (18.6)	52.6 (7.2-137.5)	0.5026
	Myfortic 1.44 g	249 (90)	52.1 (18.0)	51.1 (6.8-111.2)	
Month 9	EVR 1.5 mg	234 (84)	55.9 (18.3)	54.8 (5.8-131.5)	0.0210
	EVR 3.0 mg	248 (89)	54.1 (18.8)	53.2 (8.9-145.5)	0.4584
	Myfortic 1.44 g	250 (90)	52.5 (17.5)	50.7 (6.3-127.8)	
Month 12	EVR 1.5 mg	245 (88)	56.3 (20.1)	55.3 (4.6-140.9)	0.0573
	EVR 3.0 mg	244 (87)	55.0 (19.8)	53.8 (8.7-124.0)	0.4795
	Myfortic 1.44 g	248 (90)	54.4 (26.4)	50.8 (6.8-366.4)	
Month 12 TEP	EVR 1.5 mg	274 (99)	53.4 (21.9)	52.7 (4.4-155.9)	0.0412
	EVR 3.0 mg	278 (99)	50.2 (22.5)	50.8 (5.5-124.0)	0.8390
	Myfortic 1.44 g	272 (98)	49.6 (19.3)	48.8 (4.7-105.8)	
Month 12 SEP	EVR 1.5 mg	276 (99)	54.5 (20.7)	55.0 (4.4-140.9)	0.0324
	EVR 3.0 mg	279 (100)	51.7 (21.4)	52.0 (5.2-124.0)	0.9941
	Myfortic 1.44 g	277 (100)	52.6 (26.1)	50.2 (5.1-366.4)	

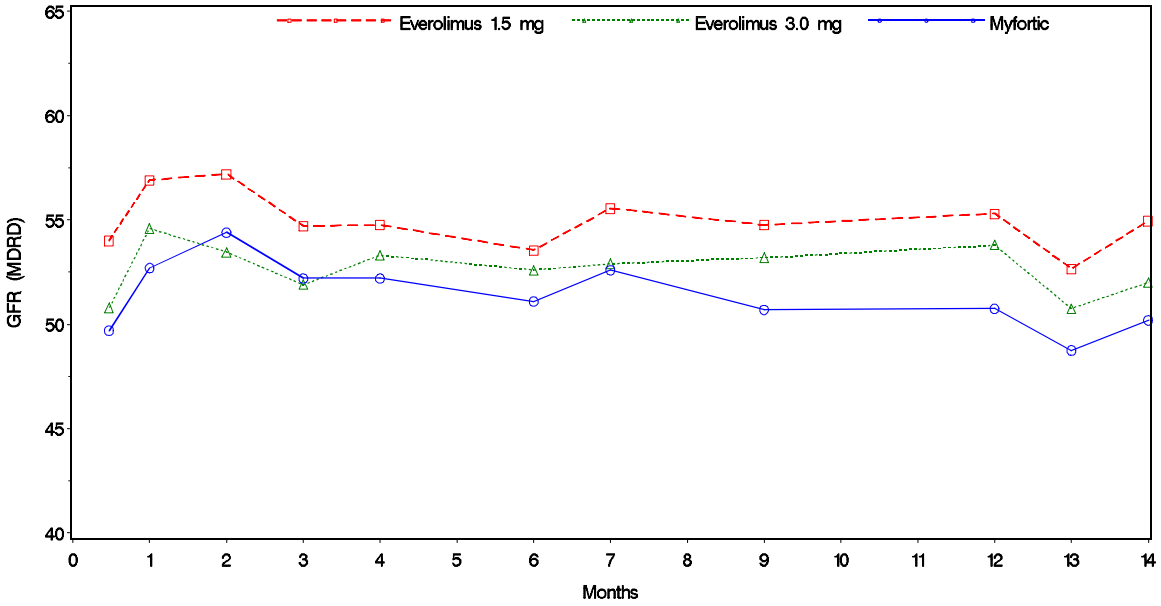
*GFR given in mL/min/1.73 m^2; ** Wilcoxon Rank Sum test; *** % is based on ITT populations; EVR = everolimus; TEP = treatment endpoint; SEP = study endpoint; No differences noted between treatment groups at baseline and Days 1, 3 and 7 (data omitted from table).

Figure 2: Mean Calculated GFR (MDRD) (ITT Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit. Month 14 represents the Month 12 study endpoint consisting of the last post-baseline observation up to and including the Month 12 visit.

Figure 3: Median Calculated GFR (MDRD) (ITT Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit. Month 14 represents the Month 12 study endpoint consisting of the last post-baseline observation up to and including the Month 12 visit.

Table 11: Calculated GFR* (MDRD) by Clinically Relevant Categories (ITT Population)

	GFR Range		GFR Range (Month 12), n (%)				
Treatment	(Month 1)	n (%)	<30	≥30 to <50	50 to <55	55 to <60	≥60
Everolimus 1.5 mg	<30	19 (8)	6 (32)	8 (42)	0 (0)	0 (0)	5 (26)
	≥30 to <50	61 (26)	6 (10)	30 (49)	11 (18)	6 (10)	8 (13)
	50 to <55	24 (10)	0 (0)	10 (42)	7 (29)	3 (13)	4 (17)
	55 to <60	23 (10)	0 (0)	6 (26)	3 (13)	6 (26)	8 (35)
	≥60	105 (45)	2 (2)	14 (13)	12 (11)	11 (11)	66 (63)
	Total	232	14 (6)	68 (29)	33 (14)	26 (11)	91 (39)
Myfortic	<30	17 (7)	6 (35)	9 (53)	1 (6)	0 (0)	1 (6)
	≥30 to <50	76 (33)	4 (5)	45 (59)	14 (18)	6 (8)	7 (9)
	50 to <55	32 (14)	1 (3)	16 (50)	4 (13)	7 (22)	4 (13)
	55 to <60	21 (9)	1 (5)	7 (33)	4 (19)	3 (14)	6 (29)
	≥60	87 (37)	1 (1)	17 (20)	6 (7)	11 (13)	52 (60)
	Total	233	13 (6)	94 (40)	29 (12)	27 (12)	70 (30)

*GFR given in mL/min/1.73 m^2

3.2.3.2 Proteinuria

Proteinuria was assessed by evaluating the ratio of spot urine protein (measured in grams) to creatinine (measured in grams) based on an estimate of an average 24-hour excretion and were collected by central laboratory. Analyses of the data indicate that the UP/UC ratios distributions are skewed due to extreme outlying values such that the means were greater than the medians. To account for the lack of symmetry in the data, data were analyzed by comparing medians at each visit window between treatment groups and using the nonparametric Wilcoxon rank-sum test to test for treatment differences.

Table 12 includes the summary statistics for UP/UC ratios across selected visit windows in the safety on-treatment population. Of note, ratios in the everolimus 3.0 mg group are not presented in Table 12; however, the everolimus 3.0 mg group consistently performed worse than everolimus 1.5 mg when compared to Myfortic. These data are also plotted in Figure 4 illustrating that the medians ratios in everolimus 1.5 mg group were consistently higher than those of Myfortic throughout the 12-month study period including the Month 12 TEP, shown as Month 13 in the figure. The everolimus 1.5 mg and Myfortic arms were statistically significantly different at all visit windows after Day 14 except at Month 4.

Beginning at Month 1 until Month 12, there were 6, 12 and 8 patients in the everolimus 1.5 mg, everolimus 3.0 mg and Myfortic arms, respectively, who had at least one UP/UC ratio greater than 5 g/g. When these values were excluded from the analysis, the results did not change significantly.

Note also, that there was an increasing number of missing data as study follow-up time increased. At Day 14, about 90% of data were collected compared to only about 70% at Month 12, as shown in Table 12. Additionally, the rates of premature treatment discontinuation at 12-month post-transplant were 30% (83/277), 34.1% (95/279) and 21.7% (60/277) in the everolimus

1.5 mg, everolimus 3.0 mg and Myfortic groups respectively resulting in a p-value=0.03 (everolimus 1.5 mg v. Myfortic) and p=0.001 (everolimus 3.0 mg v. Myfortic). The differential premature treatment discontinuation rates could lead to biased results in the on-treatment analyses.

Table 12: UP/UC Ratios* by Visit Window (Safety On-treatment Population)

Visit Window	Treatment Group	n (%)**	Mean (SD)	Median (Range)	p-values***
Day 14	EVR 1.5 mg	241 (89)	0.60 (1.29)	0.33 (0.06-17.05)	0.1881
	Myfortic	244 (90)	0.62 (1.28)	0.29 (0.08-13.04)	
Month 1	EVR 1.5 mg	246 (91)	0.43 (0.76)	0.26 (0.06-8.51)	0.0025
	Myfortic	244 (90)	0.40 (0.85)	0.20 (0.06-9.51)	
Month 3	EVR 1.5 mg	219 (81)	0.28 (0.42)	0.17 (0.02-3.90)	0.0338
	Myfortic	224 (83)	0.27 (0.50)	0.13 (0.05-4.19)	
Month 6	EVR 1.5 mg	188 (69)	0.25 (0.43)	0.15 (0.00-4.96)	0.0292
	Myfortic	207 (77)	0.25 (0.50)	0.12 (0.00-4.65)	
Month 9	EVR 1.5 mg	188 (69)	0.25 (0.30)	0.15 (0.00-2.24)	<0.0001
	Myfortic	198 (73)	0.22 (0.42)	0.11 (0.03-3.88)	
Month 12	EVR 1.5 mg	183 (68)	0.31 (0.59)	0.15 (0.03-6.15)	<0.0001
	Myfortic	192 (71)	0.27 (0.61)	0.11 (0.00-5.12)	
Month 12 TEP	EVR 1.5 mg	271	0.70 (3.64)	0.21 (0.03-58.00)	<0.0001
	Myfortic	270	0.49 (1.23)	0.12 (0.00-10.39)	

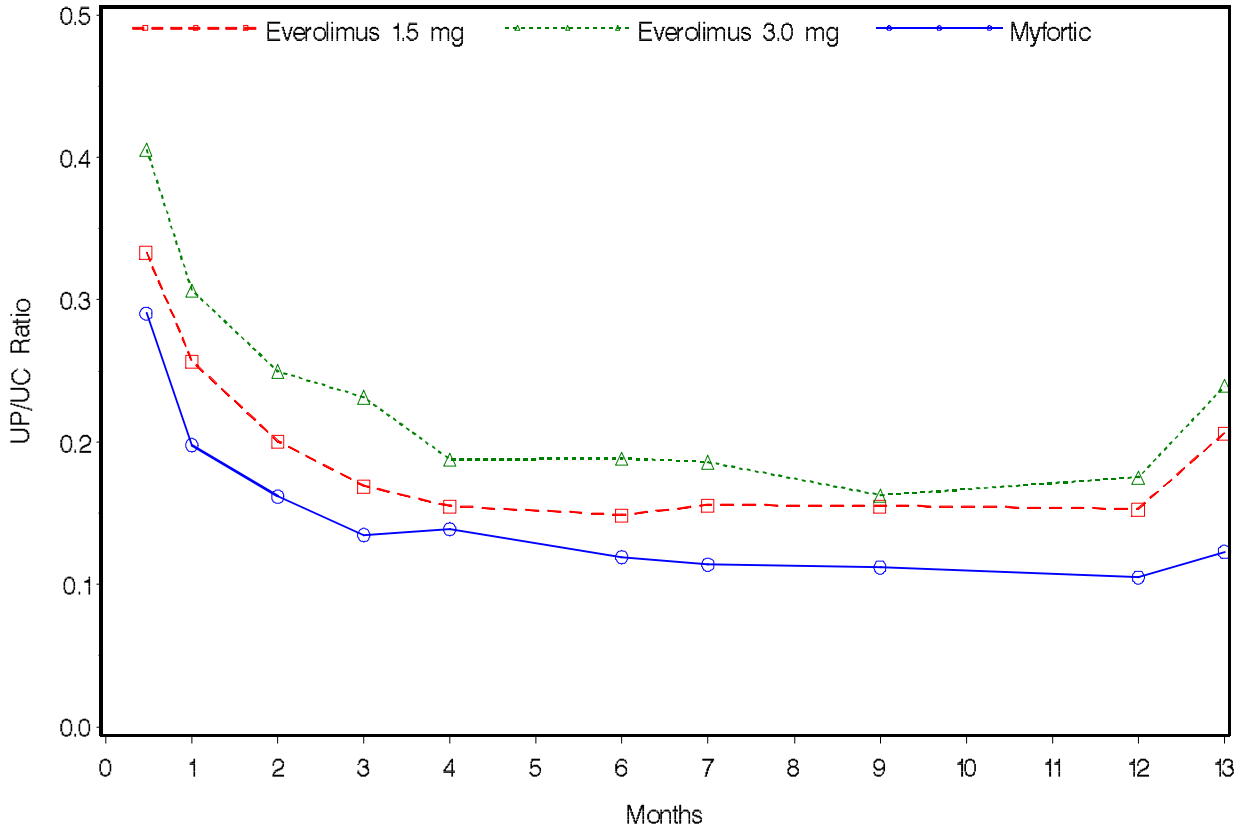
*In gram/gram unit; **% relative to Month 12 TEP; ***Wilcoxon rank-sum-test; TEP=treatment endpoint (imputation by on-treatment LOCF); EVR=Everolimus; No differences noted between treatment groups at baseline, Days 1, 3, and 7 (data omitted from table)

The applicant performed analysis of the UP/UC ratio using a linear piecewise regression model with a breakpoint at Day 28 for each treatment arm. The applicant showed that when outliers were removed, there were no statistical differences of parameters estimates between treatment groups. However, the piecewise model was not reliable because the residuals were skewed suggesting deviation from normality.

Reviewer's Comment: *The skewed residuals in the linear piecewise regression model for UP/UC ratio imply that the model is not appropriate for this data because it assumes normality. See section 5.1 for more about the asymmetry of the UP/UC ratio data.*

UP/UC ratios falling into clinically relevant categories at each visit window are presented in Table 13. These data suggest that there were more UP/UC ratios that were less than 2.0 in the Myfortic group than the everolimus 1.5 mg group from Month 1 through Month 12. This result is consistent with what was previously observed when assessing median ratios.

Figure 4: Median UP/UC Ratio in gram/gram (Safety On-treatment Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

The proportion of patients, by treatment group, with a UP/UC ratio falling into a specific clinically-relevant range at Month 12 by Month 1 ratio are presented in Table 14 below. These results suggest the following:

- At Month 1, 95% of patients in the everolimus 1.5 mg group had a UP/UC ratio < 1 compared to 96% of patients in the Myfortic group. Of these patients, 40% and 56% in the everolimus 1.5 mg and Myfortic groups, respectively, had UP/UC ratios below 0.2 at Month 1 and 76% and 85% of these maintained that level at Month 12.
- There were many more patients in the everolimus 1.5 mg arm than Myfortic (99 versus 74, respectively) who had UP/UC ratios in the >0.2 to < 1 category at Month 1. Of patients in this category, 92% (everolimus 1.5 mg) and 94% (Myfortic) either maintained these ratios or had improved ratios falling into the ≤ 0.2 category at Month 12.
- At Month 12, compared to everolimus 1.5, there were proportionately more patients in the Myfortic group with lower UP/UC ratios.

Table 13: UP/UC Ratio* by Visit Window (Safety On-treatment Population)

			UP/UC Ratio, n (%)**			
Visit Window	Treatment Group	n (%)***	Normal ≤0.2	Mild >0.2 - <1	Sub-nephrotic 1 - < 3	Nephrotic ≥3
Baseline	EVR 1.5 mg	145 (54)	4 (3)	56 (39)	54 (37)	31 (21)
	Myfortic	134 (50)	3 (2)	45 (34)	58 (43)	28 (21)
Day 1	EVR 1.5 mg	235 (87)	7 (3)	136 (58)	70 (30)	22 (9)
	Myfortic	235 (87)	2 (1)	125 (53)	85 (36)	23 (10)
Day 14	EVR 1.5 mg	241 (89)	61 (25)	151 (63)	25 (10)	4 (2)
	Myfortic	244 (90)	78 (32)	138 (57)	21 (10)	7 (3)
Month 1	EVR 1.5 mg	246 (91)	87 (35)	143 (58)	14 (6)	2 (1)
	Myfortic	244 (90)	125 (51)	104 (43)	12 (5)	3 (1)
Month 3	EVR 1.5 mg	219 (81)	132 (60)	79 (36)	7 (3)	1 (0)
	Myfortic	224 (83)	146 (65)	70 (31)	5 (2)	3 (1)
Month 6	EVR 1.5 mg	188 (69)	124 (66)	60 (32)	3 (2)	1 (1)
	Myfortic	207 (77)	147 (71)	52 (25)	6 (3)	2 (1)
Month 9	EVR 1.5 mg	188 (69)	120 (64)	62 (33)	6 (3)	.
	Myfortic	198 (73)	149 (75)	42 (21)	6 (3)	1 (1)
Month 12	EVR 1.5 mg	183 (68)	110 (60)	65 (36)	6 (3)	2 (1)
	Myfortic	192 (71)	143 (74)	40 (21)	5 (3)	4 (2)
Month 12 TEP	EVR 1.5 mg	271	135 (50)	111 (41)	17 (6)	8 (3)
	Myfortic	270	175 (65)	67 (25)	17 (6)	11 (4)

*In gram/gram unit ;** % of row total; *** % of Month 12 TEP; EVR=everolimus

Note that Table 14 consists of data from patients who had both Months 1 and 12 UP/UC measurements, representing only about 70% of the Month 12 TEP sample size.

Table 14: Categorized UP/UC Ratios* (Months 1 and 12)

	UP/UC Range		UP/UC Range (Month 12), n (%)**			
Treatment	(Month 1)	n (%)	Normal ≤0.2	Mild >0.2 to <1	Sub-nephrotic 1 to <3	Nephrotic ≥3
Everolimus 1.5 mg	≤0.2	72 (40)	55 (76)	16 (22)	1 (1)	0 (0)
	>0.2 to <1	99 (55)	52 (52)	40 (40)	5 (5)	2 (2)
	1 to <3	9 (5)	2 (22)	7 (78)	0 (0)	0 (0)
	≥3	1 (1)	0 (0)	1 (100)	0 (0)	0 (0)
	Total	*181*	*109 (60)*	*64 (35)*	*6 (3)*	*2 (1)*
Myfortic	≤0.2	104 (56)	88 (85)	14 (13)	2 (2)	0 (0)
	>0.2 to <1	74 (40)	47 (64)	22 (30)	2 (3)	3 (4)
	1 to <3	7 (4)	2 (29)	3 (43)	1 (14)	1 (14)
	≥3	2 (1)	1 (50)	1 (50)	0 (0)	0 (0)
	Total	*187*	*138 (74)*	*40 (21)*	*5 (3)*	*4 (2)*

*In gram/gram unit; % of Month 12 TEP

There were 25 (9%) reports of proteinuria as an adverse event (AE) in the everolimus 1.5 mg group and 20 (7%) reports in the Myfortic group. Two proteinuria AEs in the everolimus 1.5 mg group and 1 in the Myfortic group were reported as serious AEs. Two proteinuria AEs in the

everolimus 1.5 mg group and none in the Myfortic group lead to drug discontinuation. For more details about proteinuria as an adverse event, please refer to the clinical review.

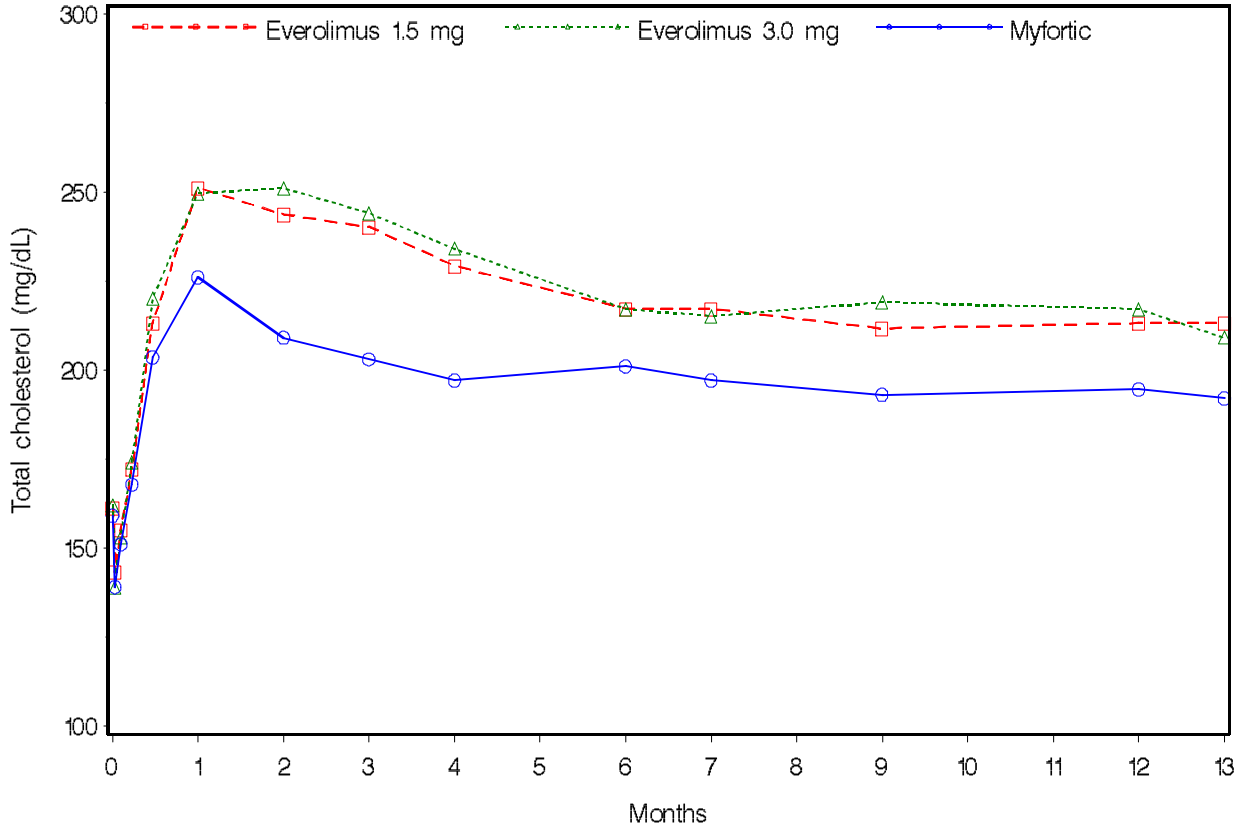
3.2.3.3 Lipids

Lipids were assessed in the safety on-treatment population focusing on the following clinical parameters: total cholesterol, triglycerides, LDL, HDL and cholesterol-HDL ratio. The data presented in the figures below are given in mg/dL units. Since the distributions of the lipid measurements at each visit window are skewed, medians were plotted and the treatment groups were compared using the Wilcoxon rank-sum test.

3.2.3.3.1 Total Cholesterol

As illustrated in Figure 5 median total cholesterol was consistently higher in both everolimus groups compared to the Myfortic group and statistically significant differences were found from Month 1 post-transplant through Month 12 TEP (Month 13 in the figure). There were no statistically significant differences among treatment arms at baseline.

Figure 5: Median Total Cholesterol (Safety On-treatment Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

Note: There was a decreasing trend in the number of observations collected per treatment arm post-baseline across time. The number of measurements per treatment arm was within 89%-93% at Day 14 and Month 1, 69%-81% at Month 6 and only 63%-75% at Month 12.

The total cholesterol levels were also analyzed using the following categorization in mg/dL units: <200, 200 to <240 and ≥240. The results are presented in Table 15, showing that across visit windows, a larger percentage of patients in the Myfortic group had total cholesterol <200 mg/dL (considered a clinically-relevant lower cut-off for hypercholesterolemia) compared to the everolimus 1.5 mg group. This finding is consistent with Myfortic having a lower median than everolimus 1.5 mg through the 12-month follow-up period (Figure 5).

Table 15: Categorized Total Cholesterol* (Safety On-treatment Population)

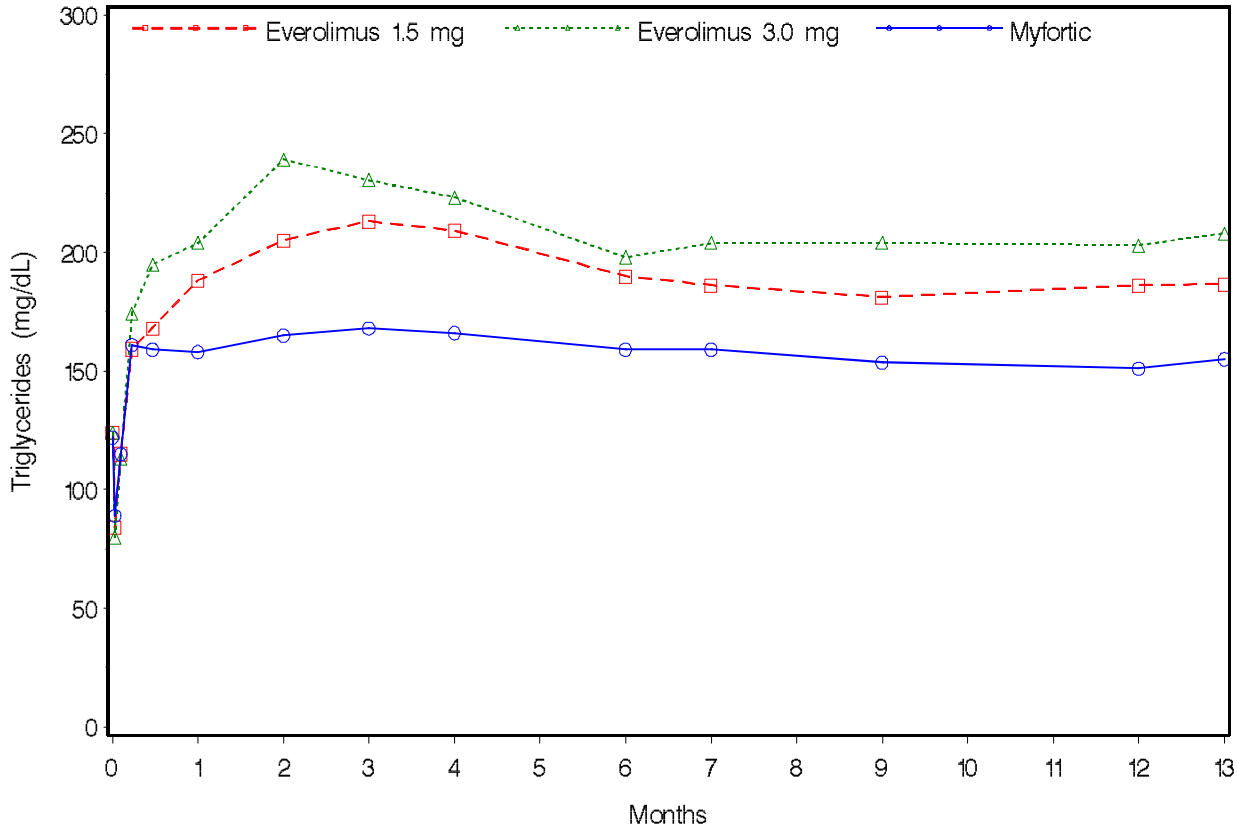
			Total Cholesterol, n (%)**		
Visit Window	**Treatment Group**	**n (%)*** **	**< 200**	**200 to < 240**	**> 240**
Baseline	EVR 1.5 mg	259 (95)	208 (80.3)	37 (14.3)	14 (5.4)
	Myfortic	261 (96)	206 (78.9)	44 (16.9)	11 (4.2)
Day 1	EVR 1.5 mg	243 (89)	224 (92.2)	12 (4.9)	7 (2.9)
	Myfortic	245 (90)	227 (92.7)	15 (6.1)	3 (1.2)
Day 14	EVR 1.5 mg	253 (92)	98 (38.8)	77 (30.4)	78 (30.8)
	Myfortic	250 (92)	115 (46.0)	83 (33.2)	52 (20.8)
Month 1	EVR 1.5 mg	247 (90)	58 (23.5)	45 (18.2)	144 (58.3)
	Myfortic	252 (93)	82 (32.5)	68 (27.0)	102 (40.5)
Month 3	EVR 1.5 mg	224 (82)	60 (26.8)	51 (22.8)	113 (50.4)
	Myfortic	233 (86)	104 (44.6)	76 (32.6)	53 (22.8)
Month 6	EVR 1.5 mg	196 (72)	66 (33.7)	53 (27.0)	77 (39.3)
	Myfortic	221 (81)	108 (48.9)	65 (29.4)	48 (21.7)
Month 9	EVR 1.5 mg	194 (71)	87 (44.8)	43 (22.2)	64 (33.0)
	Myfortic	210 (77)	117 (55.7)	55 (26.2)	38 (18.1)
Month 12	EVR 1.5 mg	188 (69)	75 (39.9)	64 (34.0)	49 (26.1)
	Myfortic	204 (75)	111 (54.4)	61 (29.9)	32 (15.7)
Month 12 TEP	EVR 1.5 mg	274	109 (39.8)	86 (31.4)	79 (28.8)
	Myfortic	272	158 (58.1)	73 (26.8)	41 (15.1)

*In mg/dL; ** % of row total; *** % of Month 12 TEP; EVR=Everolimus

3.2.3.3.2 Triglycerides

Median triglycerides, shown in Figure 6, were consistently higher in both everolimus arms compared to Myfortic from Month 1 through Month 12 follow-up. Differences between the everolimus 1.5 mg and Myfortic arms were statistically significant at Month 1 and onwards, including the Month 12 TEP. No differences were noted in baseline triglyceride values.

Figure 6: Median Triglycerides (Safety On-treatment Population)



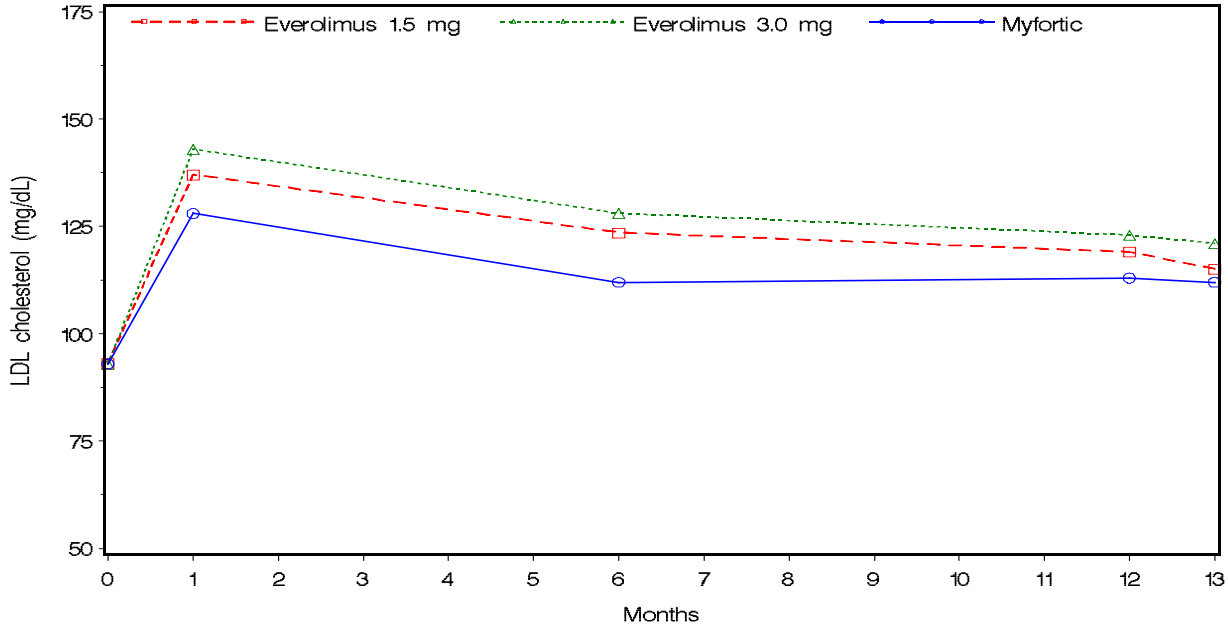
Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

3.2.3.3.3 LDL, HDL and Total Cholesterol to HDL Ratio

LDL, HDL and total cholesterol to HDL ratio were assessed at Baseline, Months 1, 6 and 12. Few measurements were obtained at other study visits and were therefore excluded in the analysis. At baseline, no differences were noted among treatment arms. Post-baseline LDL and HDL levels were statistically significantly different between everolimus 1.5 mg and Myfortic at Month 1 (p-value=0.0086) and at Month 12 TEP (p-value=0.0158) for LDL (Figure 7) and Month 6 (p-value=0.0013) and at Month 12 TEP (p-value=0.0002) for HDL (Figure 8). Median post-baseline cholesterol to HDL ratio in the everolimus 1.5 mg arm were greater than in Myfortic except at Month 6, though treatment differences were not statistically significant (Figure 9).

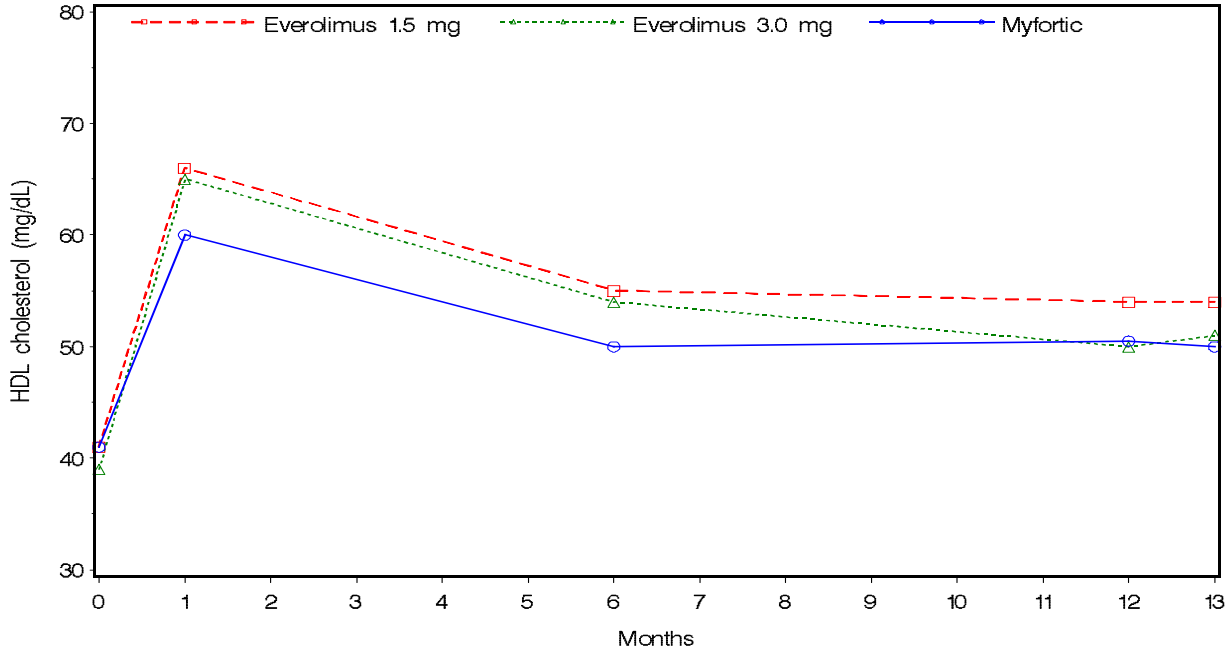
Post-baseline LDL and HDL levels were statistically significantly different between everolimus 3.0 mg and Myfortic at Month 1 (p-value=0.0080), Month 6 (p-value=0.0183), Month 12 (p-value=0.0022) and at Month 12 TEP (p-value=0.0012) for LDL (Figure 7) and Month 12 TEP (p-value=0.0201) for HDL (Figure 8). Median post-baseline cholesterol to HDL ratios in the everolimus 3.0 mg arm were greater than in Myfortic and differences were statistically significant at Month 12 (p-value=0.0013) and Month 12 TEP (p-value=0.0142) (Figure 9).

Figure 7: Median LDL (Safety On-treatment Population)



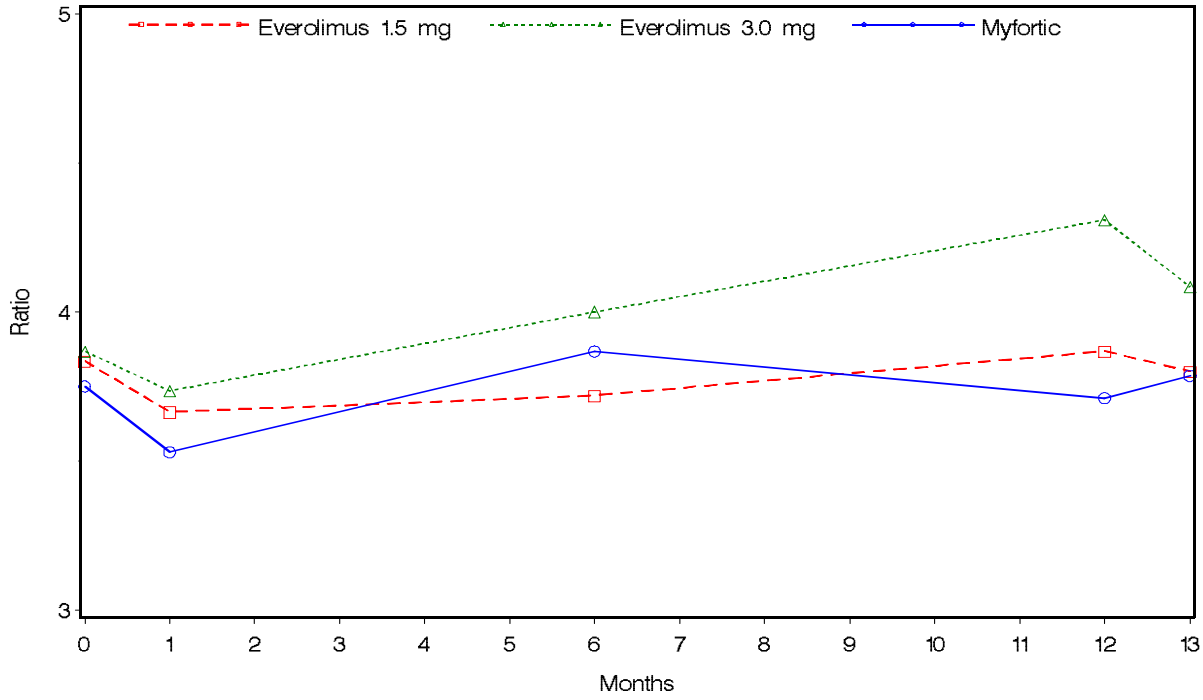
Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

Figure 8: Median HDL (Safety On-treatment Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

Figure 9: Median Total Cholesterol/HDL Ratio (Safety On-treatment Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

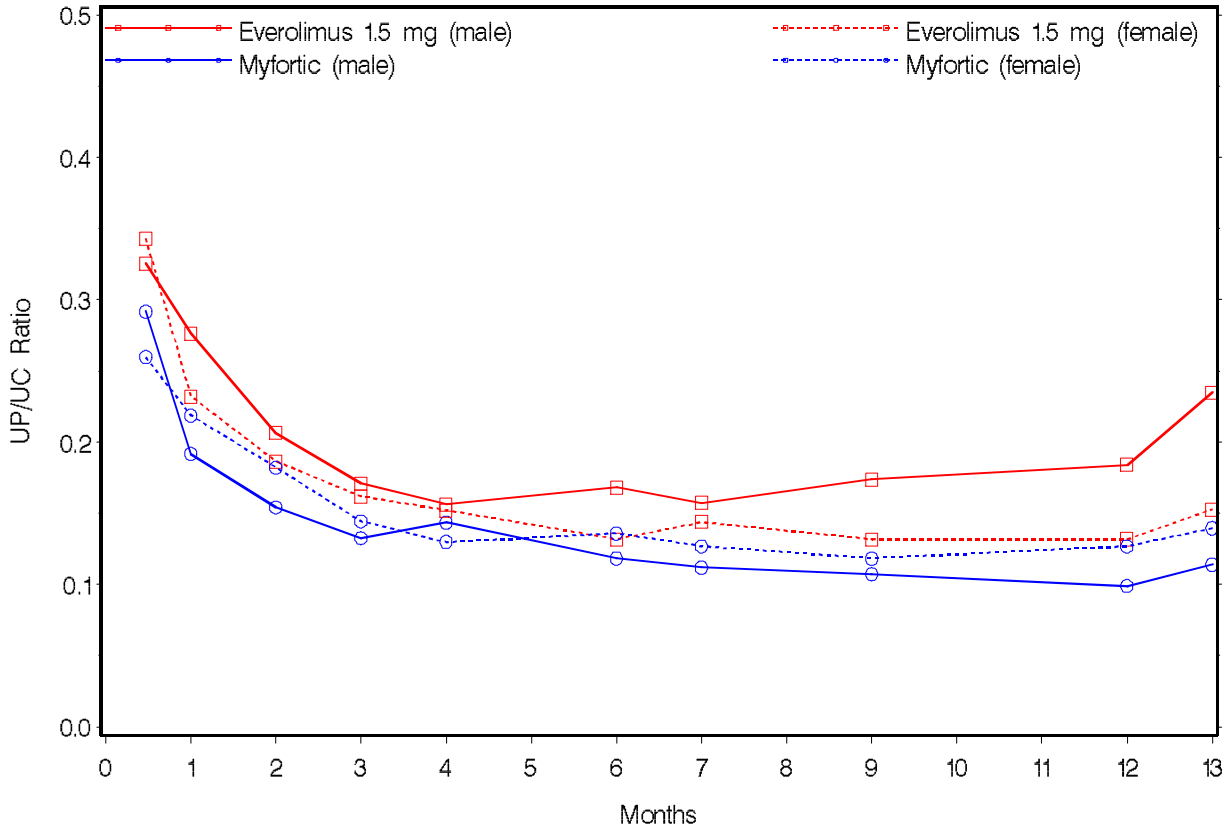
4.1 Gender, Race and Age

4.1.1 Proteinuria by Gender

Among male patients, median UP/UC ratios were consistently higher in the everolimus 1.5 mg compared to the Myfortic arm. The differences between treatment arms were found to be statistically different at all time points except at Baseline, Days 7, 14 and at Month 4, as shown in Figure 10. These differences between the everolimus 1.5 mg and Myfortic arms were not observed among female patients. Thus, it appears that the differences between the everolimus 1.5 versus Myfortic arms in the overall population may have been driven by the differences among male patients. *Note: These multiple comparisons are unadjusted.*

There were no apparent trends between the everolimus 3.0 mg and Myfortic groups by gender (results not presented). This could be related to the fact that more patients discontinued treatment early in the everolimus 3.0 mg group compared to Myfortic, which may have led to a bias in reporting.

Figure 10: Median UP/UC Ratio by Gender (Safety On-treatment Population)



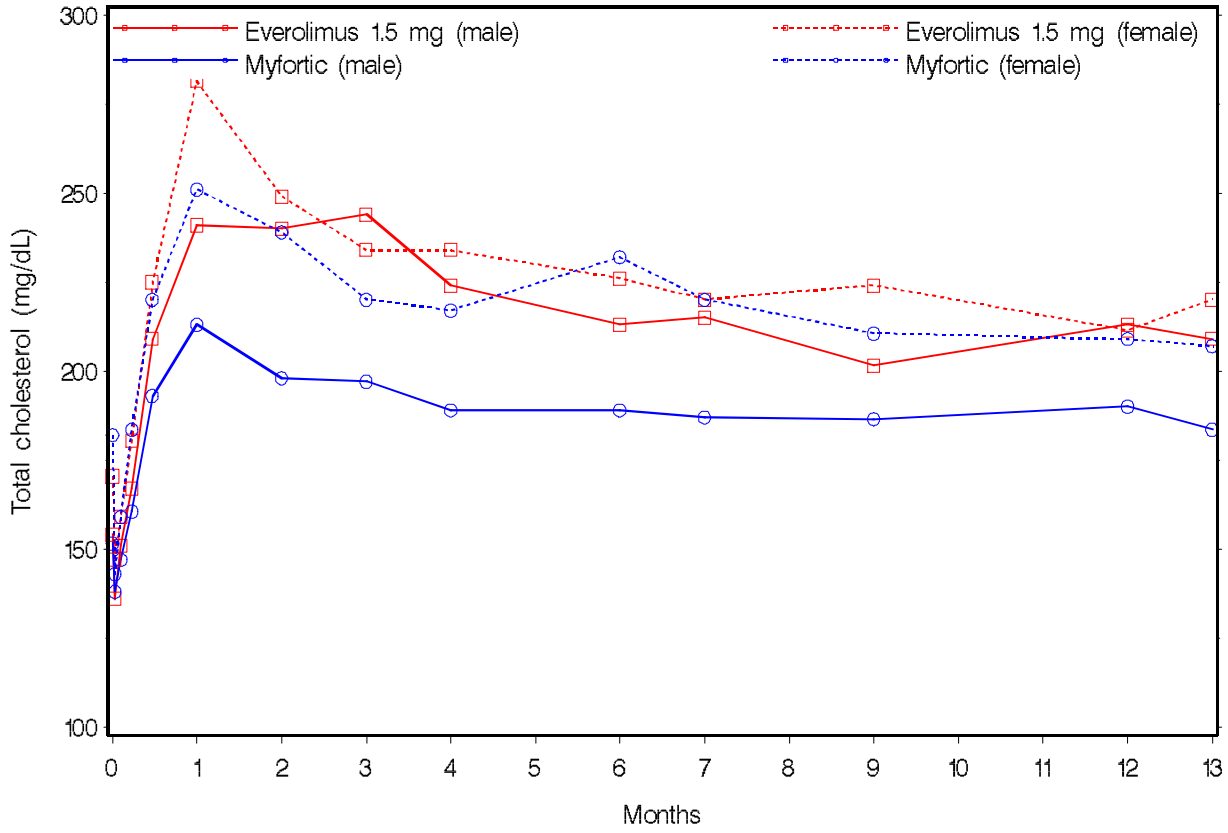
Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

4.1.2 Lipids by Gender

Among male patients (solid lines in Figure 11), median total cholesterol levels were consistently higher in the everolimus 1.5 mg arm compared to the Myfortic arm and the treatment arms were significantly different at all visit windows except at Baseline and Day 1. Among female patients (dashed lines in Figure 11), differences between the everolimus 1.5 mg and Myfortic arms were only observed at Months 1 (p-value 0.0449), 2 (p-value 0.0215) and 3 (p-value 0.0420) though total cholesterol values among females in both treatment groups were generally higher than among males. Thus, while females in both treatment groups had higher total cholesterol values overall, it appears that the statistical differences between the everolimus 1.5 versus Myfortic arms may have been driven by significant differences between treatment groups in the subset of males in the study. *Note: These multiple comparisons are unadjusted.*

Figure 11 also suggests that the subgroup of males in the Myfortic arm generally have lower total cholesterol levels compared to all other subgroups i.e. males and females in everolimus 1.5 mg and females in Myfortic.

Figure 11: Median Total Cholesterol by Gender (Safety On-treatment Population)



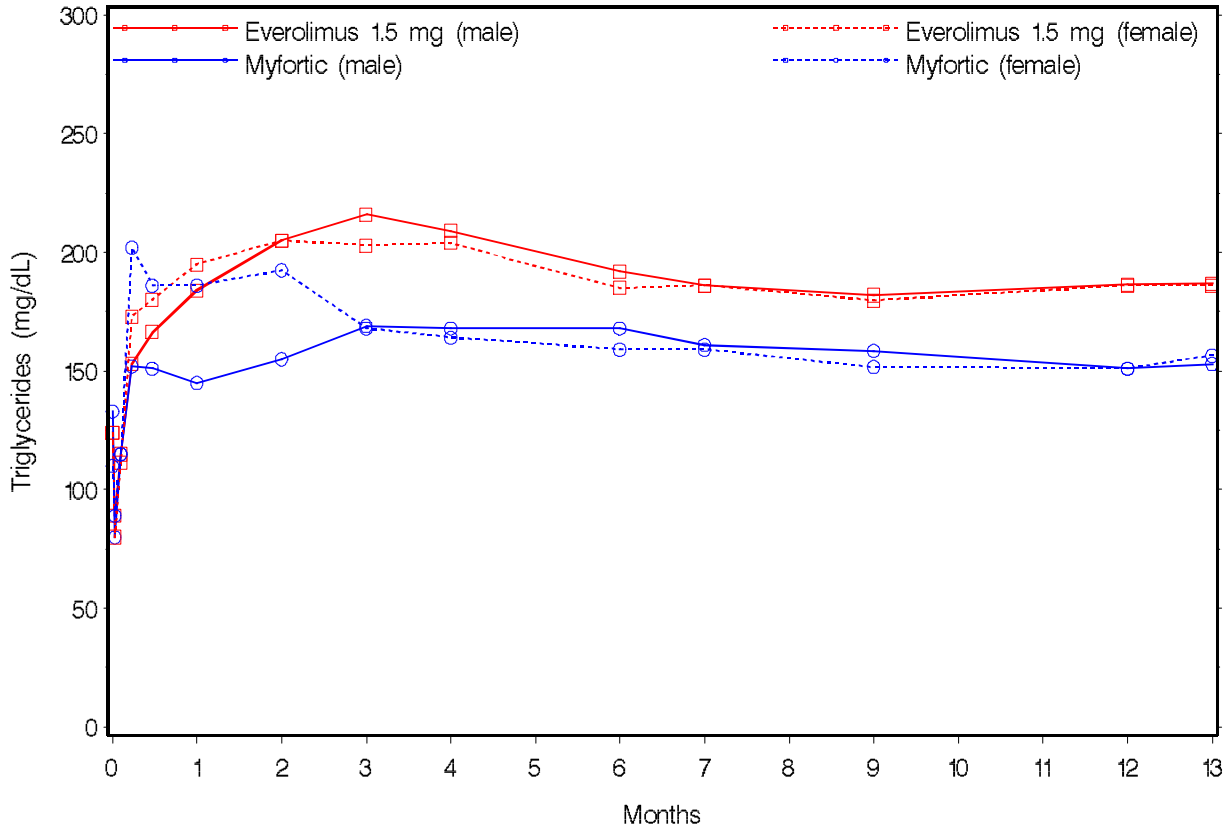
Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

Median triglyceride levels among males (solid lines in Figure 12) in the everolimus 1.5 mg arm were consistently higher than among males in the Myfortic arm. The differences between treatment arms were statistically significant at all visit windows except at Baseline and Days 1, 3, 7 and 14. A similar trend was found among females (dashed lines in Figure 12), with statistically significant differences noted between the everolimus 1.5 mg and Myfortic arms at Months 3, 6, 7, 9 and 12, including the Month 12 TEP. Thus, for triglyceride levels, it appears that the differences between the everolimus 1.5 versus Myfortic arms in the overall population are also seen when looking at each subgroup by gender. This suggests a lack of interaction between gender and treatment in the analysis of triglycerides.

There were no notable gender differences or trends observed in other subgroup analyses of LDL, HDL and total cholesterol to HDL ratio.

Subgroup analyses by race (Blacks versus non-Blacks) and age (50 years or older versus less than 50 years old) did not show any significant results with respect to all safety parameters that were assessed in this review.

Figure 12: Median Triglycerides by Gender (Safety On-treatment Population)



Note: Month 13 represents the Month 12 treatment endpoint consisting of the last post-baseline on-treatment observation up to and including the Month 12 visit.

4.2 Other Special/Subgroup Populations

Subgroup analyses according to baseline diabetic status did not show any significant differences between treatment groups with respect to all safety parameters that were assessed in this review.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The visit windows described in the study A2309 SAP and used by the applicant for assessments and analyses were inconsistent with what the protocol specified. The widths of the visit windows were not uniform because the visits were not equally spaced. It is unclear how the results of the analyses would change if this were not the case.

There were disproportionate and statistically significant rates of premature treatment discontinuation in study A2309: 30% and 34% in everolimus 1.5 mg and 3.0 mg, respectively, versus 22% in Myfortic (p=0.03 everolimus 1.5 mg v. Myfortic, p=0.001 everolimus 3.0 mg v. Myfortic, Fisher's exact test). The rates were driven mainly by adverse event-related discontinuations. There were also an increasing number of missing observations across time: up

to 20% per arm for GFR and up to 30% per arm for UP/UC ratio and lipids at Month 12. These issues can bias results from the analysis of all clinical parameters assessed in this review. The resulting estimates may not be accurate or precise due to a reduced sample size and increased variability. Furthermore, the large number of missing observations requires a large number of imputations using LOCF which may also lead to biased sensitivity analyses. Additionally, the rates of reported AEs may be biased in favor of the everolimus groups since AEs were collected only up to 8 days following end of treatment.

GFR was the only clinical parameter assessed in this review that had a pre-specified analysis in the protocol. Assessment of GFR for non-inferiority of everolimus treatment arms compared to the active control Myfortic was carried out using a t-test and supplemented with several imputation methods for missing observations as sensitivity analyses. The primary imputation method was LOCF which may have biased the sensitivity analysis because there were observations that were imputed at Month 12 from visit windows as early as Day 1. Although pre-specified in the protocol, the non-inferiority margin of 8 ml/min/1.73 min^2 was not justified as it was not derived from historical data. The other assessments in this review (proteinuria and lipids) were requested by the clinical reviewers.

The marginal distributions of measurements (GFR, UP/UC ratio, and lipid parameters) at particular visit windows were skewed (asymmetric) due to extreme outlying observations. This implied that the marginal distributions were not normal and analyses based on statistical tests (e.g. t-test) and models that assume normality (e.g. ANCOVA and linear mixed effects models) were inappropriate. This issue was most pronounced for the UP/UC ratio. This could have been due to the large amount of missing assessments, particularly at later study time points.

The subgroup analyses in section 4 showed that the differences between the everolimus 1.5 mg and Myfortic arms in the overall population for UP/UC ratio and total cholesterol may have been driven by differences between groups among male study patients. These differences did not seem to exist among females in the study. For the UP/UC ratio, the differences between the everolimus 3.0 mg and Myfortic did not appear to be gender related. For total cholesterol, it was observed in section 4.2 that the subgroup of male study patients in the Myfortic arm generally had lower total cholesterol levels compared to all other subgroups, i.e. males and females, receiving everolimus and females receiving Myfortic. The differences between everolimus 1.5 mg and Myfortic in the overall population for triglycerides were also seen in each gender subgroup with no apparent interaction.

5.2 Conclusions and Recommendations

Study A2309 demonstrated that calculated 12-month GFR, using the modification of diet in renal disease (MDRD) formula, was similar between both everolimus groups and the Myfortic group (Table 1). Various sensitivity analyses, modeling and imputation methods for missing values resulted in similar results in 12-month GFR across treatment groups. Analyses of GFR trends found that the median GFR levels in the everolimus 1.5 mg group were numerically higher than those of Myfortic across most study visit windows but the treatment groups were not statistically significantly different at all time points.

In both everolimus groups, there were statistically significant findings for proteinuria [as measured by urinary protein to urinary creatinine (UP/UC) ratio]. Specifically, the median UP/UC ratios for both everolimus arms were numerically higher than those of Myfortic at all measured time points after Day 3 (Days 7, 14 and Months 1 to 12). Differences in UP/UC ratio between the everolimus 1.5 mg and Myfortic groups were statistically significant at all time points from Month 1 until Month 12 (all p-values<0.05), except at Month 4 (p-value=0.11). Similarly, statistically significant differences between everolimus 3.0 mg and Myfortic were observed at Day 14 until Month 12 (all p-values<0.05). The differences in the median UP/UC ratios between the everolimus 1.5 mg arm and Myfortic, which appeared to increase beginning at 6-months post-transplant and continued through Month 12, appeared to be driven by differences between treatment groups in the subgroup of male study patients.

Total cholesterol levels were statistically significant and the medians were numerically higher in both everolimus arms compared to Myfortic from Month 1 post-transplant through Month 12 follow-up (p-values<0.05). From Month 1 onwards, the median total cholesterol values in both everolimus arms remained above 200 mg/dL, which is considered the clinical lower limit for hypercholesterolemia. Additionally, triglycerides levels were statistically significant and the medians were numerically higher in both everolimus arms compared to Myfortic from Month 1 through Month 12 follow-up (p-values<0.05). The median total cholesterol differences between the everolimus treatment arms and Myfortic appeared to be driven by differences between treatment groups in the subgroup of male study patients.

LDL and HDL levels were statistically significantly different between everolimus 1.5 mg and Myfortic at Month 1 (p-value=0.0086) and at Month 12 treatment endpoint (p-value=0.0158) for LDL and Month 6 (p-value=0.0013) and at Month 12 treatment endpoint (p-value 0.0002) for HDL. Note: The Month 12 treatment endpoint was defined as the last post-baseline *on-treatment* observation up to and including the scheduled Month 12 visit. Median post-baseline cholesterol to HDL ratios in the everolimus 1.5 mg arm were greater than in Myfortic except at Month 6, though treatment differences were not statistically significant. LDL levels were statistically significant at Month 1 (p-value=0.0080), Month 6 (p-value=0.0183), Month 12 (p-value=0.0022) and Month 12 treatment endpoint (p-value=0.0012) and the medians were numerically higher in the everolimus 3.0 mg group compared to Myfortic. Median HDL levels in the everolimus 3.0 group were numerically higher than in Myfortic but statistically significant differences between the two groups were only observed at the Month 12 treatment endpoint (p-value=0.0201).

SIGNATURES/DISTRIBUTION LIST (Optional)

Primary Statistical Reviewer: John Stephen Yap, PhD
Date: December 10, 2009

Concurring Reviewer(s):

Statistical Team Leader (Acting): LaRee Tracy, PhD

Biometrics Deputy Division Director (Acting): Aloka Chakravarty, PhD

Application Type/Number	Submission Type/Number	Submitter Name	Product Name
NDA-21560	ORIG-1	NOVARTIS PHARMACEUTICALS CORP	CERTICAN (EVEROLIMUS) TABLETS

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

/s/

JOHN S YAP
12/10/2009

LAREE A TRACY
12/10/2009

ALOKA G CHAKRAVARTY
12/10/2009

STATISTICAL NDA REVIEW AND EVALUATION

Medical Division: **Division of Special Pathogen and Transplant Products**

Biometrics Division: **Division of Biometrics IV**

NDA:	21-560/SN007 and SN009 (submission received 4/7/09)
NAME OF DRUG:	Certican (everolimus)
INDICATION(S):	Prophylaxis of organ rejection in kidney transplantation
APPLICANT:	Novartis Pharmaceutical Corporation
DOCUMENTS REVIEWED:	Briefing Package for FTF MTG to discuss re-submission
STATISTICAL REVIEWER:	LaRee Tracy, PhD
STATISTICAL TEAM LEADER:	Karen Higgins, ScD
BIOMETRICS DIVISION DIRECTOR:	Mohammad Huque, PhD
CLINICAL REVIEWER:	Ergun Velidedeoglu, MD
PROJECT MANAGER:	Jackie Smith, MA

Background:

Novartis received two approvable letters (10/20/2003 and 8/27/2004) for the kidney indication. In both reviews, the Division found Certican to be efficacious in preventing acute rejection in kidney transplantation; however, significant renal toxicity was found to be associated with Certican regimen. An acceptable risk/benefit ratio was not shown in previous studies of Certican in de novo kidney transplantation. Subsequently, the sponsor designed and conducted a new phase III clinical trial evaluating a concentration controlled regimen of Certican given with reduced dose cyclosporine. This submission includes a briefing package for a face-to-face meeting (request by sponsor on 3/20/2009) between the sponsor and the Division to discuss the 12 month results of the phase III study (study A2309) as well as to obtained Division feedback in so far as the final proposal for resubmission of NDA 21-560 for kidney transplantation. Previous reviews of study A2309 can be located under IND 52,003.

Brief summary of 12 month findings (Study A2309):

A total of 833 patients provided data for this 12-month analysis and were randomized between the three treatment groups (RAD 1.5 mg dose n= 277, RAD 3.0 mg dose n= 279, Myfortic 1.44 g n=277). The number of patients in the RAD 1.5 mg, RAD 3.0 mg, and Myfortic arms who discontinued study medication by month 12 was 83 (30%), 95 (34%) and 60 (22%) respectively. The number of patients in the RAD 1.5 mg, RAD 3.0 mg, and Myfortic arms who discontinued study by month 12 was 38 (14%), 33 (12%) and 28 (10%) respectively.

Incidence of the composite endpoint (BPAR, Graft loss, death or lost to follow-up) at 12 months was 25.3%, 21.5% and 24.2% in the RAD 1.5 mg, RAD 3.0 mg, and Myfortic arms respectively. The difference (as reported by the sponsor) between the RAD 1.5 mg and Myfortic arms was 1.1%, 95% CI (-6.1%, 8.3%). The difference (as reported by the sponsor) between the RAD 3.0 mg and Myfortic arms was -2.7%, 95% CI (-9.7%, 4.3%).

Incidence of graft loss or death was 18%, 21% and 15% in the RAD 1.5 mg, RAD 3.0 mg, and Myfortic arms respectively.

The 12-month mean GFR (ml/min/1.73m2) in the RAD 1.5 mg, RAD 3.0 mg, and Myfortic arms was 54.55, 51.29 and 52.18 respectively resulting in a difference of 2.37, 95% CI (-1.7, 6.4) between the RAD 1.5 and Myfortic arms and of -0.89, 95% CI (-5.0, 3.2) between the RAD 3.0 and Myfortic arms.

The submission included a list of questions for the Division. The following is a summary of the Division's responses.

Question 1:

Does the Division agree that the various analyses of the primary efficacy endpoint (treated BPAR, graft loss, death or loss to follow-up) demonstrate that Certican in a regimen with reduced dose Neoral is non-inferior to the myfortic and standard dose Neoral group?

FDA Response: Preliminary results provided in the meeting briefing document suggest that the noninferiority objective was achieved; however, a detailed review and assessment of the completed A2309 study is necessary in order to conclude that one or both Certican regimens is non-inferior to the active control. Additionally, either in the resubmission or as a separate submission to the IND, please provide a detailed quantitative justification for the chosen 10% noninferiority margin used in study A2309.

Reviewer's Comment: During the FTF meeting, the Division recommended that the sponsor use an approach to justify the NI similar to that used to for the AEB071 study. This justification should consider how the efficacy of CellCept is equivalent to that of Myfortic.

Question 2:
Does the Division agree that the various analyses of the renal function endpoint appropriately show non-inferiority and that these results are acceptable for NDA resubmission?

FDA Response: The Division does not consider non-inferiority approaches appropriate for evaluation of renal function in kidney transplantation. Additionally, the Division did not agree to the chosen renal function non-inferiority margin. Ultimately evaluation of the renal function including different key components such as proteinuria will be a review issue, data on efficacy endpoints will be assessed to determine if a favorable benefit to risk ratio was achieved. Although GFR is an important component of renal function, proteinuria is also another component and is an important marker of kidney injury and a predictor of graft survival. While on the surface the results based on GFR may seem acceptable for a resubmission, the review will closely assess whether the reduction in CNI nephrotoxicity is offset by a different and equally concerning type of nephrotoxicity such as proteinuria. In table 11-9 of the summary report for study 2309 there seems to be a trend towards progressive increase of proteinuria in both the 3mg and the 1.5mg Certican arms compared to the Myfortic arm starting at month 6. Since we only have data up to 12 months it is not possible to say if this differential increase in proteinuria will continue over time but it is known that this is a class effect of M-TOR inhibitors and may require treatment with ACE inhibitors in some cases.

Reviewer's Comment: Simple analysis of renal data by comparing mean values at specific time points is a limited approach. Therefore, during the FTF meeting, statistics suggested that the sponsor consider additional methods to assess renal function including mixed effects modeling and time to event analyses. Additionally, ANCOVA including baseline characteristics of interest should be considered.

Question 3:

Does the Division agree that study results show a reasonable compliance with everolimus and cyclosporine drug levels to support safe dose recommendations?

FDA Response: We noticed that the proportion of patients whose CsA concentrations were within the target ranges was declining as a function of time in the everolimus arms. In other words, during Months 3 and above, for the majority of patients, the CsA concentrations were actually above the target ranges for both everolimus treatment arms. In comparison the Myfortic group of patients had a higher proportion of patients whose CsA concentrations were within the target range throughout the study.

We recommend that you perform exposure-response analyses as a function of both CsA and everolimus concentrations in the resubmission, as you had previously performed in NDAs 21-560 and 21-628 (three-dimensional plots to describe the relationship of CsA and everolimus concentrations vs. effectiveness and safety endpoints).

Question 4:
Does the Division agree that preliminary safety results suggest an acceptable profile for recommending use of Certican in kidney transplantation?

FDA Response: In this 1:1:1 randomized study the total number of deaths are 9, 7 and 6 in the 3mg, 1.5mg and the Myfortic arms respectively. In the Myfortic arm one of the deaths is due to a traffic accident and one more is listed as caused by "injury, poisoning and procedural complications". Before having the narratives of these cases it is not possible to say to what extent these deaths are related to the treatment regimens. In the Certican heart transplant study 2310 which utilized similar treatment arms and regimens as in this study, the 3mg arm was terminated early due to three times as many deaths compared to the control arm. We see a similar trend in this study as well in the 3mg Certican arm, and if the same trend exists in the 1.5 mg arm remains to be seen. Drug discontinuations due to adverse events in the Certican arms are approximately twice as many as in the control arm (18%, 20% vs 9%). The summary data suggests that there may be an advantage in favor of the Certican arms regarding the incidence of leucopenia, CMV and BK virus infections and neoplasms but a disadvantage regarding proteinuria, hypercholesterolemia, peripheral edema, wound problems, lymphocele and mouth ulcers. Wound complications requiring surgery were seen in 19 and 24 patients in the 1.5 mg and 3 mg arms vs. 10 patients in the Myfortic arm. Although the preliminary results do not suggest an acceptable safety profile for any of the Certican arms ultimately this will be a review issue. The Division requests additional information about the cases with interstitial lung disease and FSGS if available.

Question 5:

Do the Division statisticians have any further comments on the revised Statistical Analysis Plan? Does the Division recommend analysis beyond those in the SAP version 2.0 to further characterize the safety profile of Certican?

FDA Response: The division would like to see the detailed narratives of the deaths, lost to follow-up cases, drug discontinuations and detailed analyses of the cases with proteinuria, hyperlipidemia and peripheral edema including the percentage of patients with high end values. A detailed description of the methodology utilized in the assessment of UP/UC ratio and the definitions of CMV and BK virus infections will also be helpful. The Division also requests a grading system be utilized for cases with peripheral edema if this was included in the CRFs.

Question 6:
Do the Division medical reviewers want a similar evaluation of wound healing and related complications to that presented in the recent publication by Tiong HY, et. al. (Transplantation 2009)

FDA Response: The analysis method used in the Tiong paper is not very helpful in assessing the cases with wound dehiscences including superficial and fascial dehiscences and eviscerations. If Novartis prefers to do a similar analysis in addition to the standard analysis of wound related complications this will be considered as supportive.

Question 7:
Will the Division accept a Clinical Overview (eCTD Module 2.5) providing summary information without an accompanying separate Summary of Clinical Efficacy and Summary of Clinical Safety?

FDA Response: This approach is acceptable.

Additional Comment:
Please plan to submit analysis data sets with this NDA submission. These analysis data sets should contain both source and derived variables and allow for easy recreation of analyses related to primary and safety objectives.

Note: The FTF meeting was held on 5/6/2009. Minutes of this meeting are available in DARRTs.

Linked Applications	Submission Type/Number	Sponsor Name	Drug Name / Subject
NDA 21560	GI 1		CERTICAN (EVEROLIMUS) TABLETS
NDA 21560	GI 1		CERTICAN (EVEROLIMUS) TABLETS

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

/s/

LAREE A TRACY
08/18/2009

KAREN M HIGGINS
08/19/2009

STATISTICAL NDA REVIEW AND EVALUATION

Medical Division: **Division of Special Pathogen and Transplant Products**

Biometrics Division: **Division of Biometrics IV**

NDA:	21-560 (dated 8/21/08 and 10/23/08)
NAME OF DRUG:	Certican® (everolimus)
INDICATION(S):	Prophylaxis of organ rejection in kidney transplantation
APPLICANT:	Novartis Pharmaceuticals Corporation
DOCUMENTS REVIEWED:	Type C Meeting Request (8/21/08) and Briefing Packet (10/23/08)
STATISTICAL REVIEWER:	LaRee Tracy, MA
STATISTICAL TEAM LEADER:	Karen Higgins, ScD
BIOMETRICS DIVISION DIRECTOR:	Mohammad Huque, PhD
CLINICAL REVIEWER:	Erun Velidedeoglu, MD
PROJECT MANAGER:	Jackie Smith, MA

Summary:
The original NDA 21-560 for Certican tablets in kidney transplantation was submitted on 12/19/2002, which included two pivotal studies (B201 and B251) in de novo kidney transplantation. Both studies evaluated a regimen of fixed dose Certican in combination with full dose cyclosporine compared to MMF with standard doses of cyclosporine. Due to an unfavorable safety profile shown in the data provided, the Division issued an approvable letter in October 2003. In February 2004, the sponsor submitted a major amendment to the NDA including two additional studies (A2306 and A2307) both of which were uncontrolled studies in de novo kidney transplantation. The Division issued a second approvable letter due to a lack of findings to support a safety and effective regimen in kidney transplantation. Novartis subsequently agreed to conduct a large, prospective, Phase 3 clinical trial (study A2309) in de novo kidney transplantation to evaluate a concentration-controlled regimen of Certican in combination with concentration-controlled cyclosporine compared to a regimen of Myfortic with standard dose cyclosporine. Protocol A2309 is entitled, "***A 24-month, multicenter, randomized, open-label, non-inferiority study of efficacy and safety comparing concentration-controlled Certican in two doses (1.5 and 3.0 mg/day starting doses) with reduced Neoral versus 1.44 g Myfortic with standard dose Neoral in de novo renal transplant recipients***". Previous reviews of this protocol and analyses plan can be located in DAARTS under IND 52,003 SNs 339, 362, 374, 421, 424, 433 and SDN 724.

On 8/22/08, Novartis requested a meeting with the Division to discuss new proposals and timelines for resubmission of the Certican NDA for kidney transplantation. The briefing book with questions was sent to the Division on 10/23/08 for the 11/24/08 teleconference.

Summary
The purpose of the meeting was to obtain feedback regarding the proposed contents of a future resubmission for NDA 21-560, which would be based on 12-month results from study A2309.

The following is a summary of the statistical/clinical related questions and the Division's preliminary responses (sent to sponsor on 11/20/08):

Use of everolimus therapeutic drug monitoring (TDM)
The sponsor asked to know the Division's expectations for use of TDM for everolimus with reduced dose cyclosporine in kidney transplantation and if the Division required TDM to improve the safety profile associated with everolimus.

The Division's response noted that the use of everolimus with reduced dose cyclosporine in organ transplantation should maintain adequate protection against rejection while providing an adequate safety profile. We further noted that we expected adequate compliance by study investigators with the TDM regimens for everolimus and cyclosporine.

Refer to the preliminary comments sent to the sponsor for the full response.

Ongoing study A2309
The sponsor as if in view of the anticipated safety profile of everolimus would demonstration of non-inferiority with respect to the pre-specified efficacy endpoint (BPAR, graft loss, death or lost to follow-up) provide adequate evidence of benefit.

The Division responded that this is a review issue and the Division's decision will be based on a thorough risk/benefit assessment of the proposed regimen for the specified population. Demonstration of NI may not be sufficient to overcome a serious safety finding and conversely if the study demonstrates a benefit with respect to graft loss or death that could offset the risk of a rare serious adverse event.

Class-related safety considerations in kidney transplantation

The sponsor asked for additional guidance on current safety related considerations to support the evaluation of the complete response.

The Division provided a list of current safety concerns that the sponsor should assess in the review of the study results (the list can be found in the preliminary responses). Further, the Division suggested that a detailed analyses of these events, including incidence, time to event, time to resolution, and severity.

NDA Safety Update to provided side-by-side comparisons between studies A2309, B201 and B251

The sponsor if the approach to summarize, in a side-by-side fashion, key safety data from studies A2309, B201 and B251 were acceptable.

The Division noted that the proposals are reasonable (see preliminary response for the full response).

Benefit/risk assessment
The sponsor noted that they will evaluate the overall safety and efficacy of Certican in Study A2309 to the original NDA studies B201 and B251. Additionally, they will compare the benefit/risk profile of the TDM regimen of everolimus with reduced dose of cyclosporine to the overall efficacy and safety profile of the MPA active control group.

The Division noted that they agree with the proposal but cautions that cross-study comparisons are viewed as supportive only to the primary analyses of each confirmatory trial alone.

SAS Datasets and Transfer Programs
(b) (4)
.

The Division noted that they prefer that these files and datasets are resubmitted to allow for a more efficient and organized review. The proposals for what they sponsor plans to submit are acceptable otherwise.

Following receipt of the Division's preliminary responses, the sponsor responded that a face-to-face meeting was unnecessary but that a short teleconference was still needed to clarify a few additional points. The additional questions from sponsor sent on 11/21/08 along with the Division's responses provided during the 11/24/08 teleconference are as follows:

Study A2309 Statistical Analysis Plan (SAP)
The sponsor asked if the Division had any additional comments to the SAP for study A2309 (referring to IND 52,003, SN 724).

The Division replied that there are no additional comments.

Cross-study comparison from studies A2309, B201 and B251
The sponsor asked if it would be acceptable if in the cross-study comparisons that data from studies B201/B251 are re-analyzed using the visit windows and data cut-off points used in study A2309, which are different from those used in the original NDA submission.

The Division asked what the sponsor intends to accomplish with proposed cross-study comparisons. The sponsor noted that they are intended to provide a safety assessment and to demonstrate an improved safety profile compared with that from earlier studies. There will be no statistical inference drawn from these comparisons. The Division responded that side-by-side comparisons are acceptable but that study A2309 was designed to stand alone. The sponsor agreed.

Lastly, the Division reminded the sponsor that study A2309 is open-label and as such the analyses specified in the SAP will serve as primary. Any additional analyses planned, such as the discussed exposure-response analyses for specified safety events, are supportive only.

The sponsor intends to request an additional meeting prior to the NDA re-submission sometime in the first quarter of 2009.

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

/s/

LaRee Tracy
12/19/2008 02:09:24 PM
BIOMETRICS

Karen Higgins
12/19/2008 02:33:45 PM
BIOMETRICS

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21560 / class 2 complete response (kidney)
	21628 / class 2 complete response (heart)
Drug Name:	Certican (everolimus) Tablets
Indication(s):	Prophylaxis of organ rejection in allogeneic kidney and heart transplantation
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submission Date: February 27, 2004
	User Fee Date: August 27, 2004
Review Priority:	6 months
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Ruthanna C. Davi, M.S.
Concurring Reviewers:	Karen Higgins, Sc.D.
Medical Division:	Division of Special Pathogens and Immunologic Drug Products
Clinical Team:	Clinical Reviewer: Arturo Hernandez
	Clinical Team Leader: Marc Cavaillé-Coll
Project Manager:	Matt Bacho
	Andrei Nabakowski

Keywords: confounding, external control, NDA review, noninferiority

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

From a statistical perspective, randomly assigned treatment within a single study provides assurance that the treatment groups are equivalent in terms of both observed and unobserved covariates and that statistically significant differences between the two groups, if they occur, are a result of the assigned treatment, not extraneous covariates. These assurances are not provided by an externally controlled comparison such as the cross-study comparisons being proposed in this submission. Even a very small difference in the patient group and/or study characteristics may dramatically affect the probability of achieving an incorrect conclusion. Therefore, it is the opinion of this reviewer that the cross-study comparisons being proposed are not reliable and that the efficacy and safety of Certican has not been established utilizing the data in this resubmission.

1.2 Brief Overview of Clinical Studies

On December 19, 2002 the original NDAs 21560 and 21628 for the use of Certican tablet for the prophylaxis of organ rejection in adult renal and heart transplantation patients, respectively, were submitted by the sponsor. The NDAs included two pivotal renal studies (i.e., studies B201 and B251), one pivotal heart study (i.e., study B253), and two supportive renal studies (i.e., studies B156 and B157). As a result of this submission, an NDA Action letter was issued on October 20, 2003 indicating that both NDAs were approvable. The action letter stated that, "Although you have demonstrated your product to be efficacious, as studied in your clinical trials, you have yet to show a sufficiently safe regimen for everolimus when used with cyclosporine. Before these applications may be approved, you must establish a dosing regimen of everolimus and cyclosporine that is both safe and effective for the prophylaxis of organ rejection in allogeneic renal and cardiac transplant patients". The current supplement was submitted by the sponsor in response to the NDA Action letter and contains the results of two additional studies (i.e., A2306 and A2307). As these studies were originally designed to compare the 1.5 mg and 3.0 mg doses of Certican and therefore do not include an active control group, the submission is based primarily on cross-study comparisons of certain data from the A2306 and A2307 studies to certain data from the studies in the original submission.

1.3 Statistical Issues and Findings

The efficacy and safety analyses within this submission are based primarily on cross-study comparisons. (b) (4) In the opinion of this reviewer, this is inappropriate and is likely to lead to incorrect conclusions. Therefore, it is the opinion of this reviewer

that the cross-study comparisons included in this review and the sponsor's NDA submission do not represent a legitimate comparison of either the efficacy or safety of Certican relative to that of MMF since treatment was not randomly assigned.

Spurious differences between the same treatment groups in Studies B201 and B251 were observed, indicating that the studies are inherently different (see Table 2). Therefore it is the opinion of this reviewer that simplistic pooling of the results of Studies B201 and B251 is not appropriate.

2. INTRODUCTION

2.1 Overview

On December 19, 2002 the original NDAs 21560 and 21628 for the use of Certican tablet for the prophylaxis of organ rejection in adult renal and heart transplantation patients, respectively, were submitted by the sponsor. The NDAs included two pivotal renal studies (i.e., studies B201 and B251), one pivotal heart study (i.e., study B253), and two supportive renal studies (i.e., studies B156 and B157). For a thorough review of this submission please refer to the two "Statistical Review and Evaluation" documents both dated October 16, 2003.

As a result of the December 2002 submission, an NDA Action letter was issued on October 20, 2003 indicating that both NDAs were approvable. The action letter stated that, "Although you have demonstrated your product to be efficacious, as studied in your clinical trials, you have yet to show a sufficiently safe regimen for everolimus when used with cyclosporine. Before these applications may be approved, you must establish a dosing regimen of everolimus and cyclosporine that is both safe and effective for the prophylaxis of organ rejection in allogeneic renal and cardiac transplant patients".

The current supplement was submitted by the sponsor in response to the NDA Action letter and contains the results of two additional kidney studies (i.e., A2306 and A2307). As these studies are not internally controlled, the submission is based primarily on cross-study comparisons of certain data from the A2306 and A2307 studies to certain data from the studies in the original submission. The sponsor's claim is that, "The results from studies A2306 and A2307 demonstrate an improvement in the safety of Certican in combination with a reduced Neoral dosing regimen in *de novo* renal transplantation as compared to standard dose of Neoral used in the phase 3 renal studies B201, B251, and B156 while maintaining efficacy".

Of note, beginning at the time of the submission of the A2306 and A2307 study protocols and throughout the process described above, the statistical concerns regarding the use of external controls (e.g., cross-study comparisons) have been expressed to the sponsor on numerous occasions. For more details regarding the nature of these interactions please refer to Section 3.1.

This review will focus only on the kidney indication. For a review of the heart indication resubmission, please refer to the "Statistical Review and Evaluation" document authored by LaRee Tracy and dated August 2004.

2.2 Data Sources

The following data sets were submitted electronically and examined in the review of these studies.

\\Cdsesub1\n21560\N_000\2004-04-28\crt\datasets\kidney\derived\biopsy.xpt
\\Cdsesub1\n21560\N_000\2004-04-28\crt\datasets\kidney\derived\efficacy.xpt
\\Cdsesub1\n21560\N_000\2004-04-28\crt\datasets\kidney\derived\renal.xpt
\\Cdsesub1\n21560\N_000\2004-04-28\crt\datasets\kidney\derived\subset.xpt

Data sets allowing exploration of the cross-study comparisons being proposed by the sponsor in this submission were requested by the Agency on April 9, 2004. In response, the above referenced data sets were submitted by the sponsor on April 28, 2004. However, upon review, these data sets were found to have irreconcilable discrepancies in the patient identification numbers. These errors brought into question the overall accuracy of the data sets and ultimately precluded the possibility of additional analyses of either safety or efficacy outcomes by this reviewer.

3. STATISTICAL EVALUATION

The current supplement was submitted by the sponsor in response to the NDA Action letter and contains the results of two additional kidney studies (i.e., A2306 and A2307). As these studies are not internally controlled, the submission is based primarily on cross-study comparisons. The focus of the submission is a comparison of the Certican 1.5 mg, reduced Neoral (C2[1]) group of the A2306 study to the pooled MMF 2 g groups of the B201 and B251 studies. A secondary interest is the comparison of the Certican 3.0 mg reduced Neoral (C2), Simulect group of the A2307 study to the Certican 3 mg, reduced Neoral group of the B156 study as well as the comparison of the Certican 3.0 mg reduced Neoral (C2), Simulect group of the A2307 study to the Certican 3 mg, full Neoral group of B156. The sponsor's claim is that, "The results from studies A2306 and A2307 demonstrate an improvement in the safety of Certican in combination with a reduced Neoral dosing regimen in *de novo* renal transplantation as compared to standard dose of Neoral used in the phase 3 renal studies B201, B251, and B156 while maintaining efficacy". These studies are summarized in Table 1. Color-shaded areas indicate treatment groups that the sponsor is intending to compare to one another.

[1] C2 is the abbreviation for measuring Neoral whole blood concentration at 2 hours after dosing

Table 1: Summary of Renal Studies*			
Study Number	**Study Design Features**	**Treatment Groups**	**Number of Subjects**
B201	Double blind (first year), One year duration with 2 year extension	Certican 1.5 mg	194
		Certican 3.0 mg	198
		MMF 2 g	196
B251	Double blind (first year), One year duration with 2 year extension	Certican 1.5 mg	193
		Certican 3.0 mg	194
		MMF 2 g	196
A2306	Open label, One year duration	Certican 1.5 mg, reduced Neoral (C_2)	112
		Certican 3.0 mg, reduced Neoral (C_2)	125
B156	Open label, Three year duration	Certican 3.0 mg, full Neoral	53
		Certican 3.0 mg, reduced Neoral	58
A2307	Open label, One year duration	Certican 1.5 mg, reduced Neoral (C_2), Simulect	117
		Certican 3.0 mg, reduced Neoral (C_2), Simulect	139

* Yellow shading highlights the sponsor's intent to compare the Certican 1.5 mg, reduced Neoral (C_2) group with the **pooled** MMF 2 g groups. Green shading highlights the sponsor's intent to compare the Certican 3.0 mg, reduced Neoral (C_2), Simulect group with **each of the other groups**, Certican 3 mg, full Neoral and Certican 3 mg, reduced Neoral.

Reviewer Comment: *In the setting of a clinical trial, in order to balance the treatment groups in terms of both observed and unobserved patient characteristics, random treatment assignment is accepted by most as a crucial part of a study's design. Many of the concerns regarding the use of external controls are documented in the International Conference on Harmonization Harmonized Tripartite Guideline, 'Choice of Control Group and Related Issues in Clinical Trials', (also titled E10) sections 1.2, 1.3, and 2.5. In addition, numerous other FDA guidance documents and general clinical trial references also discuss the need for randomized treatment assignment and the limitations of externally controlled comparisons.*

Under limited circumstances, FDA guidance does allow for the use of externally controlled trials. When an internally controlled trial is not possible or not ethical the use of an external comparison may be considered; however, it would also be necessary for the usual course of the disease to be highly predictable, the impact of important covariates to be well-characterized, and the observed treatment effect would have to be dramatic. With this application, the disease progression relies on many covariates and is not easily predicted. In addition, Certican is administered in combination with both Neoral and corticosteroids and the contribution of each to both the efficacy and safety endpoints is not completely clear. Therefore, it is the opinion of this reviewer that the context of the cross-study comparisons proposed in this submission is not akin to the description provided in the guidance where an externally controlled comparison could be considered.

By conducting the cross-study comparisons described above, the sponsor is proposing (b) (4) *In the opinion of this reviewer, this is inappropriate in the context of this submission and is likely to lead to incorrect conclusions.*

To illustrate how a cross-study comparison can be misleading, consider a comparison of the Certican 3.0 mg group of study B201 and the Certican 3.0 mg group of study B251. That is a comparison of the same treatment at the same dose between two nearly identically designed and concurrently conducted studies. Table 2 contains this comparison for the primary and secondary efficacy endpoints for those studies.

From Table 2, it appears that the treatment administered as part of Study B251 is superior to that of Study B201 in terms of the second composite endpoint (i.e., graft loss, death, or lost to follow-up) at month 12 and 36 post-transplant as well as in terms of graft loss as a single event at all time points examined. This seems to be a clear signal that the treatment in Study B251 is superior to that of Study B201. However, since the treatments being considered in both studies are both Certican 3.0 mg, it is obvious that these statistically significant differences are leading us to incorrect conclusions. And that in fact the differences being observed between studies are not a treatment effect but are merely a result of differences in either the characteristics of the patient groups enrolled in each study (e.g., proportion of living donors, proportion of diabetic patients, etc.) or differences in the characteristics of the studies themselves (e.g., proportion of European versus U.S. study centers, unblinded trial may have fewer biopsies and therefore fewer biopsy proven acute rejections, etc.). Note that if one had been comparing *different* treatments across studies, distinguishing the possible treatment effect from that of the effect of the characteristics of each patient group or characteristics of each study would be impossible.

Table 2: Cross-Study Comparison of the Certican 3.0 mg groups in Studies B201 and B251 in terms of the Primary and Secondary Efficacy Analyses (ITT Group)			
6 Months Post-Transplant			
Endpoint	**Certican 3.0mg in B201 (N=198)**	**Certican 3.0 mg in B251 (N=194)**	**95% C.I. for Diff. in Prop.**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**52 (26.3%)**	**46 (23.7%)**	**(-11.1%, 6.1%)**
Graft Loss, Death, or Loss to Follow-up Composite	**24 (12.1%)**	**13 (6.7%)**	**(-11.5%, 0.4%)**
Biopsy-Proven Acute Rejection (single event)	36 (18.2%)	39 (20.1%)	(-5.9%, 9.8%)
Graft loss (single event)	17 (8.6%)	7 (3.6%)	(-10.1, -0.3%)*
Death (single event)	7 (3.5%)	6 (3.1%)	(-4.4%, 3.5%)
12 Months Post-Transplant			
Endpoint	**Certican 3.0 mg in B201 (N=198)**	**Certican 3.0 mg in B251 (N=194)**	**95% C.I. for Diff. in Prop.**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**60 (30.3%)**	**51 (26.3%)**	**(-12.9%, 4.9%)**
Graft Loss, Death, or Loss to Follow-up Composite	**33 (16.7%)**	**15 (7.7%)**	**(-15.6%, -2.5%)***
Biopsy-Proven Acute Rejection (single event)	39 (19.7%)	43 (22.2%)	(-5.6%, 10.6%)
Graft loss (single event)	21 (10.6%)	8 (4.1%)	(-12.0%, -1.4%)*
Death (single event)	8 (4.0%)	7 (3.6%)	(-4.6%, 3.7%)
36 Months Post-Transplant			
Endpoint	**Certican 3.0 mg in B201 (N=198)**	**Certican 3.0 mg in B251 (N=194)**	**95% C.I. for Diff. in Prop.**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**77 (38.9%)**	**66 (34.0%)**	**(-14.3%, 4.7%)**
Graft Loss, Death, or Loss to Follow-up Composite	**50 (25.3%)**	**28 (14.4%)**	**(-18.7%, -3.0%)***
Biopsy-Proven Acute Rejection (single event)	49 (24.7%)	50 (25.8%)	(-7.6%, 9.7%)
Graft loss (single event)	33 (16.7%)	15 (7.7%)	(-15.6%, -2.5%)*
Death (single event)	18 (9.1%)	13 (6.7%)	(-8.0%, 3.1%)

*Yellow shading indicates statistically significantly superior results for Certican 3.0 mg in B251 compared to Certican 3.0 mg in B201.

A computer simulation designed to estimate how often these types of incorrect conclusions would be drawn from cross-study comparisons was conducted and the results are illustrated in Figure 1. The simulation was conducted under the following assumptions.

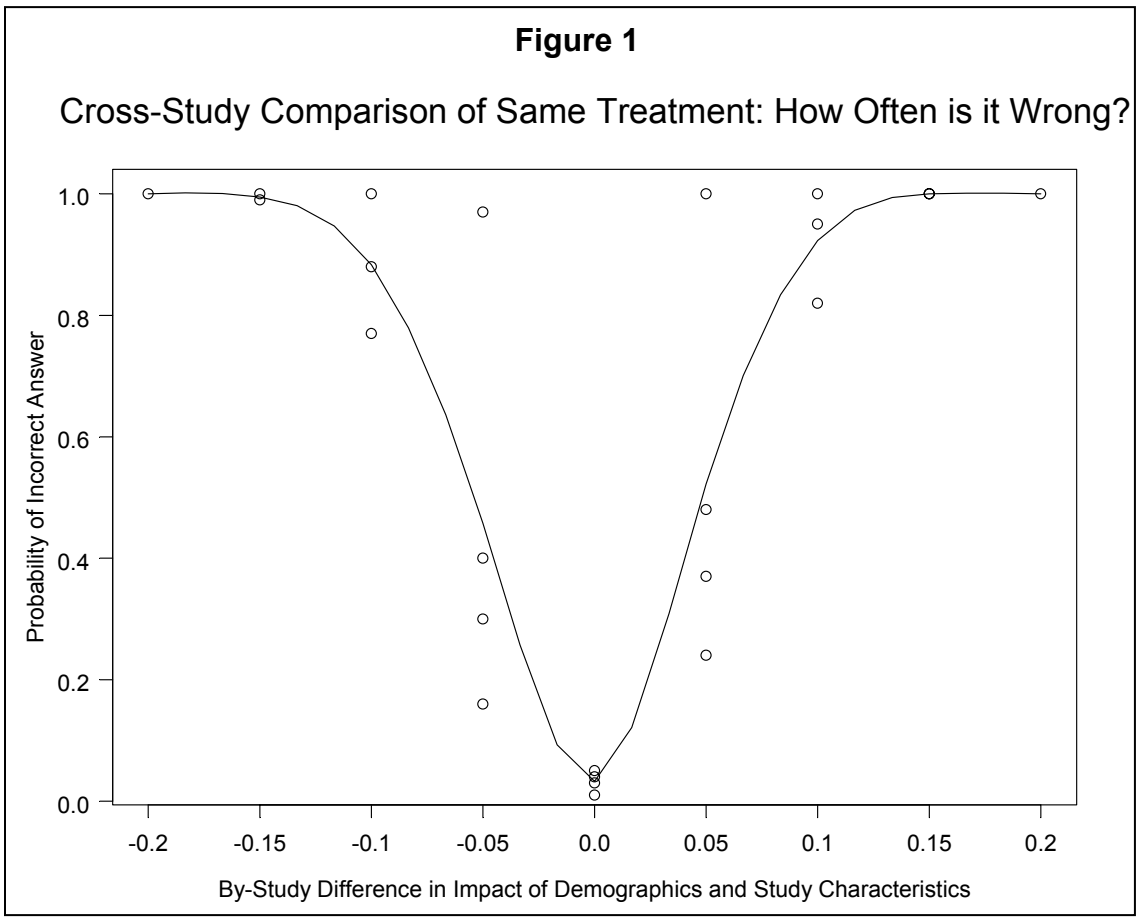
- The sample sizes of the groups being compared were set at 125 and 400. This assumption was made to mimic the proposal of the sponsor to compare one group from the A2306 study to one group from the pooled results of studies B201 and B251.
- The primary efficacy endpoint being examined is the usual composite endpoint (i.e., biopsy proven acute rejection, graft loss, death, or lost-to-follow-up) and is being analyzed using a 95% confidence interval for the difference in proportions.
- The rate of occurrence of the composite endpoint for each of the groups covered the region from 0% to 30% (in increments of 5%). This assumption was made to mimic what had been observed in recent renal transplantation clinical trials. (Note: Figure 1 includes only the range from 0% to 20% since beyond this range all probabilities were effectively equal to one.)
- As with the comparison of Certican 3.0 mg from B201 and Certican 3.0 mg from B251 above, we assumed that we were comparing identical treatments from two different studies so that any differences in the efficacy endpoint would be due to either characteristics of the patient groups and/or studies and therefore, any statistically significant differences observed between studies would be leading one to make an incorrect conclusion.

The horizontal axis in Figure 1 represents the differential impact of the characteristics of the patient groups and/or studies. For example suppose that the rate of the occurrence of the primary endpoint in the first study (due to the composition of the patient group and study characteristics) is 20% and the same such rate in the second study is 15%, then the value for the horizontal axis is 20% minus15% or 5%. The values on the vertical axis represent the probability that a statistically significant difference between groups would be detected. That is the vertical axis represents the probability of coming to the wrong conclusion. As part of the simulation, the rate of occurrence of the composite endpoint for each of the groups covered the region from 0% to 30% in increments of 5%. Each of these scenarios is represented by an open circle in Figure 1. The solid line in Figure 1 connects the mean result of the scenarios resulting in the same difference on the horizontal axis.

From Figure 1 it is apparent that the probability of making an incorrect conclusion increases as the differential impact of the characteristics of the patient groups and/or studies increases (i.e., as you move away from zero on the horizontal axis, the corresponding values on the vertical axis increase dramatically). The essential result from this simulation is that even a very small difference in the impact of patient group and/or study characteristics dramatically increases the probability of an incorrect conclusion (or in other words, the probability of a Type I error in the context of a superiority study is dramatically increased). For example suppose that the patient group and study characteristics for the studies were such that the rate of the primary event was

5% more likely in the first study than the second one. In this case, on average, the probability of an incorrect decision is over 50%.

Figure 1



Concerns regarding the use of external controls (including the cross-study comparisons described above) have been expressed to the sponsor on numerous occasions. To illustrate the nature of these interactions, excerpts from key communications regarding this topic are provided below.

- Facsimile from FDA to Novartis dated September 13, 2001 containing new protocol comments pertaining to studies A2306 and A2307

  > "…we do not believe that these phase II studies could represent adequate well-controlled studies because of the lack of valid controls. By design the studies are unable to reliably exclude an unacceptable decrease in patient or graft survival compared to approved therapy, nor would they be able to support that either dose of the test drug would have beaten a placebo. Finally, the open-label designs create a potential for bias that would impair the assessment of endpoints including but not limited to biopsy proven rejection."

- Minutes of teleconference between FDA and Novartis on November 25, 2003

"… the Division stated that there were potential problems associated with cross-study comparisons when there were differences in the observed (e.g., relative differences in enrollment of black subjects, living-related donor grafts, and the number of patients with delayed graft function) and unobserved covariates between trials. Novartis acknowledged these statements and agreed that the cross study comparisons could be systematically biased."

- Facsimile from FDA to Novartis dated February 18, 2004 containing comments pertaining to the NDA Amendment Statistical Analysis Plan

  "The issue of the validity of cross-study comparisons is a difficult one and will be addressed as part of the review of the NDA amendment. Please refer to the International Conference on Harmonization Harmonized Tripartite Guideline, 'Choice of Control Group and Related Issues in Clinical Trials', (also titled E10) sections 1.2, 1.3, and 2.5 for discussion regarding the need for concurrently controlled trials."

In an effort to address the cross-study comparison issues, the sponsor's submission provides many analyses examining the comparability of the patient populations in the different studies and the impact of various study characteristics. Overall the sponsor's conclusion from this work was that the patient populations in study A2306 and the pivotal renal studies B201 and B251 had similar demographic and background characteristics; however, there were several notable exceptions including differences in the proportions of black subjects, subject weight, proportions of subjects with living-donor grafts, and proportion of subjects with delayed graft function. The characteristics of the patients in and designs of studies A2307 and B156 were also concluded by the sponsor to be similar with several notable exceptions including the proportion of black subjects and the proportion of subjects with delayed graft function. Please refer to the medical officer review of this NDA Amendment for further discussion regarding the differences in the study characteristics and patient populations and the possible impact these differences could have on the efficacy or safety outcome. Investigation by this reviewer of the effects of covariates on either the efficacy or safety outcomes was not possible due to irreconcilable differences in the patient identification numbers in the electronic data sets provided by the sponsor.

In conclusion, from a statistical perspective, randomly assigned treatment within a single study provides assurance that the treatment groups are equivalent in terms of both observed and unobserved covariates and that statistically significant differences between the two groups, if they occur, are a result of the assigned treatment, not extraneous covariates. These assurances are not provided by an externally controlled comparison such as the cross-study comparisons being proposed. In light of the fact that even a very small difference in the impact of the patient group and/or study characteristics will dramatically increase the probability of an incorrect conclusion and the fact that certain differences in these populations and study designs have been identified it is the opinion of this reviewer that the cross-study comparisons being proposed are not reliable.

3.1 Evaluation of Efficacy

Table 3 contains the efficacy results using the sponsor's proposed cross-study comparison of the Certican 1.5 mg group from the A2306 study to the MMF groups pooled from the B201 and B251 studies. Table 4 provides the same comparison without pooling the B201 and B251 studies. Since spurious differences between the same treatment groups in Studies B201 and B251 were observed, indicating that the studies are inherently different (see Table 2), the unpooled comparison is likely more fair. However, the information in Tables 3 and 4 is included in this review for completeness only and in the opinion of this reviewer neither table contains a legitimate comparison of the efficacy of Certican with that of MMF since treatment was not randomly assigned.

Table 3: Cross-Study Comparison of the Certican 1.5 mg, reduced Neoral group in Study A2306and MMF 2 g groups pooled from Studies B201 and B251 in terms of the Primary and Secondary Efficacy Analyses (ITT Group)			
6 Months Post-Transplant			
Endpoint	**Certican 1.5mg in A2306 (N=112)**	**MMF 2 g groups pooled from B201 and B251 (N=392)**	**95% C.I. for Diff. in Prop. (Certican minus MMF)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**31/112 (27.7%)**	**109/392 (27.8%)**	**(-9.5%, 9.3%)**
Graft Loss, Death, or Loss to Follow-up Composite	**5/112 (4.5%)**	**27/392 (6.9%)**	**(-11.6%, 5.9%)***
Biopsy-Proven Acute Rejection (single event)	28/112 (25.0%)	92/392 (23.5%)	(-7.5%, 10.6%)
Graft loss (single event)	5/112 (4.5%)	22/392 (5.6%)	(-10.3%, 7.2%)*
Death (single event)	0/112 (0.0%)	5/392 (1.3%)	(-6.7%, 5.2%)*
12 Months Post-Transplant			
Endpoint	**Certican 1.5mg in A2306 (N=112)**	**MMF 2 g groups pooled from B201 and B251 (N=392)**	**95% C.I. for Diff. in Prop. (Certican minus MMF)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**31/112 (27.7%)**	**115/392 (23.3%)**	**(-11.1%, 7.8%)**
Graft Loss, Death, or Loss to Follow-up Composite	**6/112 (5.4%)**	**35/392 (8.9%)**	**(-8.0%, 2.7%)**
Biopsy-Proven Acute Rejection (single event)	29/112 (25.9%)	94/392 (24.0%)	(-7.2%, 11.1%)
Graft loss (single event)	6/112 (5.4%)	28/392 (7.1%)	(-6.7%, 3.1%)
Death (single event)	1/112 (0.9%)	9/392 (2.3%)	(-8.5%, 5.6%)*

* Calculated using exact test. All other confidence intervals calculated using the normal approximation to the binomial.

Table 4: Cross-Study Comparison of the Certican 1.5 mg, reduced Neoral group in Study A2306 and each MMF 2 g group from Studies B201 and B251 in terms of the Primary and Secondary Efficacy Analyses (ITT Group)					
6 Months Post-Transplant					
Endpoint	**Certican 1.5mg in A2306 (N=112)**	**MMF 2 g in B201 (N=196)**	**95% C.I. for Diff. in Prop. (A2306 minus B201)**	**MMF 2 g in B251 (N=196)**	**95% C.I. for Diff. in Prop. (A2306 minus B251)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**31/112 (27.7%)**	**58/196 (29.6%)**	**(-12.4%, 8.6%)**	**51/196 (26.0%)**	**(-8.7%, 12.0%)**
Graft Loss, Death, or Loss to Follow-up Composite	**5/112 (4.5%)**	**18/196 (9.2%)**	**(-15.1%, 4.6%)***	**9/196 (4.6%)**	**(-10.4%, 8.8%)***
Biopsy-Proven Acute Rejection (single event)	28/112 (25.0%)	46/196 (23.5%)	(-8.5%, 11.5%)	46/196 (23.5%)	(-8.5%, 11.5%)
Graft loss (single event)	5/112 (4.5%)	15/196 (7.7%)	(-13.5%, 6.0%)*	7/196 (3.6%)	(-6.4%, 10.3%)*
Death (single event)	0/112 (0.0%)	3/196 (1.5%)	(-10.2%, 5.5%)*	2/196 (1.0%)	(-9.7%, 6.0%)*
12 Months Post-Transplant					
Endpoint	**Certican 1.5mg in A2306 (N=112)**	**MMF 2 g in B201 (N=196)**	**95% C.I. for Diff. in Prop. (A2306 minus B201)**	**MMF 2 g in B251 (N=196)**	**95% C.I. for Diff. in Prop. (A2306 minus B251)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**31/112 (27.7%)**	**61/196 (31.1%)**	**(-14.0%, 7.1%)**	**54/196 (27.6%)**	**(-10.3%, 10.5%)**
Graft Loss, Death, or Loss to Follow-up Composite	**6/112 (5.4%)**	**22/196 (11.2%)**	**(-12.0%, 0.9%)**	**13/196 (6.6%)**	**(-6.6%, 5.1%)**
Biopsy-Proven Acute Rejection (single event)	29/112 (25.9%)	47/196 (24.0%)	(-8.2%, 12.0%)	47/196 (24.0%)	(-8.2%, 12.0%)
Graft loss (single event)	6/112 (5.4%)	18/196 (9.2%)	(-9.6%, 2.0%)	10/196 (5.1%)	(-4.9%, 5.4%)
Death (single event)	1/112 (0.9%)	5/196 (2.6%)	(-10.9%, 6.0%)*	4/196 (2.0%)	(-10.3%, 6.4%)*

* Calculated using exact test. All other confidence intervals calculated using the normal approximation to the binomial.

Table 5 contains the efficacy results using the sponsor's proposed cross-study comparison of the Certican 3.0 mg group from the A2307 study to the Certican 3.0 mg with full Neoral and the Certican 3.0 mg with reduced Neoral groups from the B156

study. As mentioned with regards to Tables 3 and 4, the information in Table 5 is included in this review for completeness only and in the opinion of this reviewer does not represent a legitimate comparison since treatment was not randomly assigned.

Table 5: Cross-Study Comparison of the Certican 3.0 mg, reduced Neoral (C2), Simulect group in Study A2307 and Certican 3.0 mg, full Neoral group in Study B156 and Certican 3.0 mg, reduced Neoral group in Study B156 in terms of the Primary and Secondary Efficacy Analyses (ITT Group)

6 Months Post-Transplant					
Endpoint	**Certican 3.0 mg in A2307 (N=139)**	**Certican 3.0 mg (full Neoral) in B156 (N=53)**	**95% C.I. for Diff. in Prop. (A2307 minus B156, full Neoral)**	**Certican 3.0 mg (reduced Neoral) in B156 (N=58)**	**95% C.I. for Diff. in Prop. (A2307 minus B156, reduced Neoral)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**27/139 (19.4%)**	**8/53 (15.1%)**	**(-7.3%, 16.0%)**	**2/58 (3.4%)**	**(7.9%, 24.1%)**
Graft Loss, Death, or Loss to Follow-up Composite	**7/139 (5.0%)**	**1/53 (1.9%)**	**(-10.5%, 18.2%)***	**1/58 (1.7%)**	**(-9.6%, 17.6%)***
Biopsy-Proven Acute Rejection (single event)	21/139 (15.1%)	8/53 (15.1%)	(-11.3%, 11.3%)	2/58 (3.4%)	(-2.2%, 26.3%)*
Graft loss (single event)	7/139 (5.0%)	1/53 (1.9%)	(-10.5%, 18.2%)*	1/58 (1.7%)	(-9.6%, 17.6%)*
Death (single event)	1/139 (0.7%)	0/53 (0.0%)	(-12.0, 12.3%)*	0/58 (0.0%)	(-11.1%, 12.3%)*
12 Months Post-Transplant					
Endpoint	**Certican 3.0 mg in A2307 (N=139)**	**Certican 3.0 mg (full Neoral) in B156 (N=53)**	**95% C.I. for Diff. in Prop. (A2307 minus B156, full Neoral)**	**Certican 3.0 mg (reduced Neoral) in B156 (N=58)**	**95% C.I. for Diff. in Prop. (A2307 minus B156, reduced Neoral)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**27/139 (19.4%)**	**15/53 (28.3%)**	**(-22.7%, 4.9%)**	**5/58 (8.6%)**	**(1.0%, 20.6%)**
Graft Loss, Death, or Loss to Follow-up Composite	**7/139 (5.0%)**	**4/53 (7.5%)**	**(-17.1%, 8.8%)***	**1/58 (1.7%)**	**(-9.6%, 17.6%)***
Biopsy-Proven Acute Rejection (single event)	22/139 (15.8%)	9/53 (17.0%)	(-12.9%, 10.6%)	4/58 (6.9%)	(-5.5%, 23.9%)*
Graft loss (single event)	7/139 (5.0%)	3/53 (5.7%)	(-15.0%, 10.3%)*	1/58 (1.7%)	(-9.6%, 17.6%)*
Death (single event)	2/139 (1.4%)	2/53 (3.8%)	(-15.7%, 7.0%)*	0/58 (0.0%)	(-10.4%, 13.9%)*

* Calculated using exact test. All other confidence intervals calculated using the normal approximation to the binomial.

3.2 Evaluation of Safety

Analyses of renal data from the original NDA submission indicated that Certican in the regimen studied including full-dose Neoral significantly and negatively impacted renal function. Therefore, the effect of Certican (even using therapeutic drug monitoring) on renal function is of concern. For this reason, the serum creatinine and creatinine clearance endpoints are being given particular attention within this review.

Table 6 and Figure 2 contain the creatinine results using the sponsor's proposed cross-study comparison of the groups in the A2306 study to the groups pooled from the B201 and B251 studies. Irreconcilable differences in the patient identification numbers in the electronic data sets precluded this reviewer from conducting the same comparison without pooling the B201 and B251 studies. The information in Table 6 and Figure 2 is included in this review for completeness only and in the opinion of this reviewer does not represent a legitimate comparison of the effect of Certican on renal function relative to that of MMF since treatment was not randomly assigned.

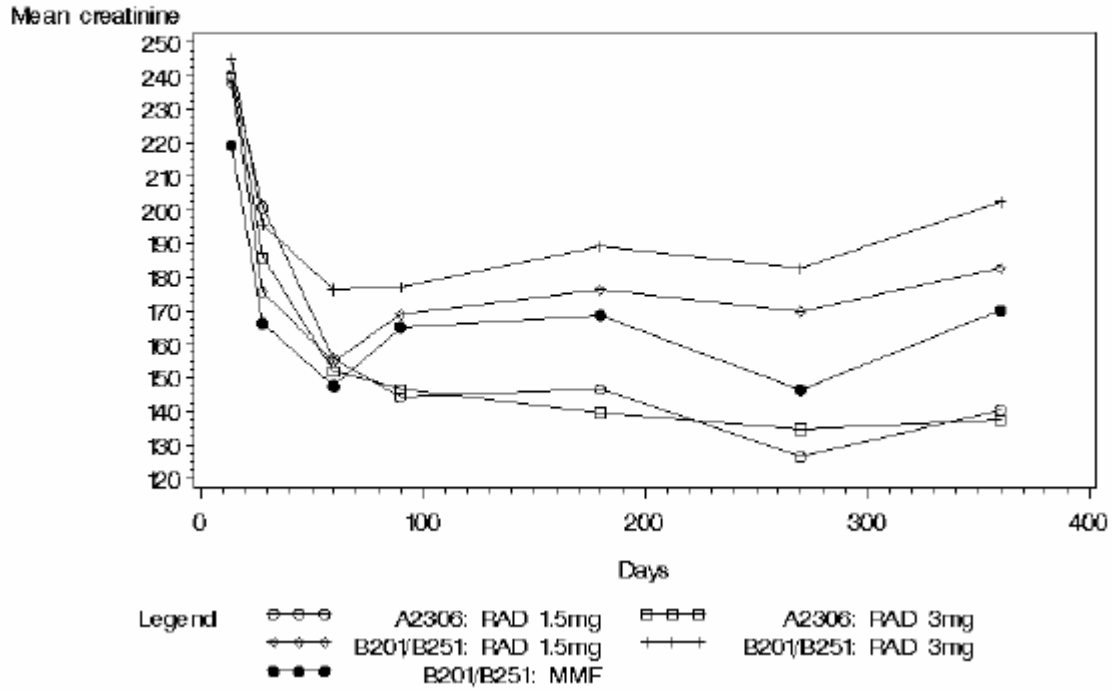
Table 6: Cross-Study Comparison of the Treatment Groups in A2306 and the Treatment Groups Pooled from Studies B201 and B251 in Terms of the Mean (Median) Creatinine

Certican 1.5 mg in A2306 (reduced Neoral) (N=105)	**Certican 1.5 mg in B201/251 (full Neoral) (N=330)**	**p-value***
140 (131)	178 (147)	<0.001
Certican 1.5 mg in A2306 (reduced Neoral) (N=105)	**MMF in B201/251 (full Neoral) (N=341)**	**p-value***
140 (131)	171 (141)	0.031
Certican 3.0 mg in A2306 (reduced Neoral) (N=111)	**Certican 3.0 mg in B201/251 (full Neoral) (N=325)**	**p-value***
137 (130)	199 (159)	<0.001
Certican 3.0 mg in A2306 (reduced Neoral) (N=111)	**MMF in B201/251 (full Neoral) (N=341)**	**p-value***
137 (130)	171 (141)	0.004

* p-value from Wilcoxon Rank-Sum test for difference between medians across studies.

Figure 2*

A2306 vs. B201/B251: Means plot of Creatinine level over the 12 month period
(Safety Population — 12—Month Analysis)



* Figure provided by sponsor in resubmission.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Please refer to the "Statistical Review and Evaluation" of the original NDA submission, dated October 16, 2003.

4.2 Other Special/Subgroup Populations

Please refer to the "Statistical Review and Evaluation" of the original NDA submission, dated October 16, 2003.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The efficacy and safety analyses within this submission are based primarily on cross-study comparisons. (b) (4)

(b) (4) In the opinion of this reviewer, this is inappropriate and is likely to lead to incorrect conclusions. Therefore, it is the opinion of this reviewer that the cross-study comparisons included in this review and the sponsor's NDA submission do not represent a legitimate comparison of either the efficacy or safety of Certican relative to that of MMF since treatment was not randomly assigned.

Spurious differences between the same treatment groups in Studies B201 and B251 were observed, indicating that the studies are inherently different (see Table 2). Therefore it is the opinion of this reviewer that simplistic pooling of the results of Studies B201 and B251 is not appropriate.

5.2 Conclusions and Recommendations

From a statistical perspective, randomly assigned treatment within a single study provides assurance that the treatment groups are equivalent in terms of both observed and unobserved covariates and that statistically significant differences between the two groups, if they occur, are a result of the assigned treatment, not extraneous covariates. These assurances are not provided by an externally controlled comparison such as the cross-study comparisons being proposed in this submission. Even a very small difference in the patient group and/or study characteristics may dramatically affect the probability of achieving an incorrect conclusion. Therefore, it is the opinion of this reviewer that the cross-study comparisons being proposed are not reliable and that the efficacy and safety of Certican has not been established utilizing the data in this resubmission.

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

/s/

Ruth Davi
8/17/04 09:22:27 AM
BIOMETRICS

Karen Higgins
8/17/04 10:00:01 AM
BIOMETRICS

Mohammad Huque
8/17/04 10:42:29 AM
BIOMETRICS

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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number:	21-560/N000 (tablets) and 21-561/N000 (dispersible tablets) for kidney indication Note: For administrative purposes, the original NDA (including both kidney and heart studies) was separated, by indication, into two NDA's. The heart indication is referenced by NDA#'s 21-628/N000 (tablets) and 21-631/N000 (dispersible tablets).
Drug Name:	Certican (Everolimus) 1.5 mg and 3.0 mg (previously SDZ RAD, Everolimus)
Indication(s):	Prophylaxis of organ rejection in allogeneic kidney and heart transplant patients
Applicant:	Novartis Pharmaceuticals Corporation
Date(s):	Submitted December 19, 2002
Review Priority:	Standard
Biometrics Division:	Division of Biometrics III
Statistical Reviewer:	Ruthanna C. Davi, M.S.
Concurring Reviewers:	Karen Higgins, Sc.D. Mohammad Huque, Ph.D.
Medical Division:	Division of Special Pathogens and Immunologic Drug Products
Clinical Team:	Medical Officer: Arturo Hernandez, M.D. Medical Team Leader: Marc Cavaille Coll, M.D.
Project Manager:	Matt Bacho

Keywords: NDA review, clinical studies, noninferiority

Table of Contents

1. EXECUTIVE SUMMARY

The sponsor has submitted the results of two controlled clinical trials (B201 and B251) in support of the efficacy of Certican for prophylaxis of organ rejection in allogeneic kidney transplant patients indication. The studies are each titled, "A three-year, double-blind, double dummy, randomized, multicenter, parallel group study of the efficacy and safety of SDZ RAD tablets versus mycophenolate mofetil as part of triple immunosuppressive therapy in de novo renal transplant recipients". The co-primary objectives of the studies were to demonstrate that the efficacies (as measured by the incidence of biopsy-proven acute allograft rejection, graft loss, death, or lost to follow-up at six months and by the incidence of graft loss, death, or lost to follow-up at 12 months) of at least one of the two oral doses of Certican are non-inferior to those of mycophenolate mofetil (MMF) in *de novo* renal transplant recipients. Studies B201 and B251 were randomized parallel-group studies. The treatment groups included 1.5 RAD, 3.0 RAD, and MMF. The double-blind studies were originally designed to last one year; however, were extended (open-label) for two additional years.

Formulation of overall conclusions regarding the efficacy of 1.5 RAD and 3.0 RAD in comparison to MMF is difficult since many by-treatment group comparisons of both the primary and secondary efficacy endpoints did not afford the luxury of consistent results. In addition sensitivity analyses conducted by this reviewer indicated that many of the primary efficacy results were not robust against the disproportionate premature treatment discontinuation rate.

The protocol prescribed co-primary efficacy analysis (of biopsy proven acute rejection, graft loss, death, or lost-to-follow up within six months) did indicate that each of the RAD groups were noninferior to MMF for both studies. However, not all of these results were robust against the effect of the disproportionate premature treatment discontinuation. When considering premature treatment discontinuation as an efficacy failure (i.e., the modified composite endpoint was biopsy proven acute rejection, graft loss, death, lost-to-follow-up, or premature treatment discontinuation) only the 1.5 RAD dose maintained non-inferiority to MMF for both of the studies. The 3.0 RAD group was not non-inferior to MMF for this endpoint in either study.

The second protocol prescribed co-primary efficacy analysis (of graft loss, death, or lost-to-follow up within 12 months) did indicate that in Study B201 the 1.5 RAD group was non-inferior to the MMF group and in Study B251 the 3.0 RAD group was non-inferior to the MMF group. However, none of these results were robust against the effect of the disproportionate premature treatment discontinuation. When considering premature treatment discontinuation as a failure (i.e., the modified composite endpoint was graft loss, death, lost-to-follow-up, or premature treatment discontinuation) neither the 1.5 RAD or 3.0 RAD groups were non-inferior to MMF in either study.

Efficacy results for the secondary analyses were not consistent across studies. Generally, the results of Study B201 seemed to indicate that the 1.5 RAD dose had more acceptable

efficacy than the 3.0 RAD dose. However, Study B251 seemed to suggest the opposite, that the 3.0 RAD dose had more acceptable efficacy. A summary of the secondary efficacy results previously discussed in this review follows.

- The 1.5 RAD group was non-inferior to the MMF group for both co-primary efficacy endpoints at all time points in Study B201.
- The 1.5 RAD group was *not* non-inferior to the MMF group for the co-primary efficacy endpoint (biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) at 36 months in Study B251. Non-inferiority for this endpoint, dose, and study was achieved for the six and 12 months time points. But results for the graft loss, death, or lost-to-follow-up co-primary composite indicated that the 1.5 RAD group was *not* non-inferior to the MMF group at any of the time points. Examination of each of the components of the composite revealed that the event of graft loss was the primary reason for these results.
- The 3.0 RAD group was non-inferior to the MMF group for the co-primary efficacy endpoint (of biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) at all time points except 36 months for Study B201. But results for the graft loss, death, or lost-to-follow-up co-primary composite indicated that the 3.0 RAD group was *not* non-inferior to the MMF group at 12 months and in fact was statistically significantly *worse* than the MMF group at 36 months. Examination of each of the components of the composite revealed that the event of graft loss was the primary reason for these results.
- The 3.0 RAD group was non-inferior to the MMF group for both the co-primary efficacy endpoints at all time points except 36 months for Study B251.

Statistically significant differences among treatment groups were observed in creatinine and creatinine clearance at numerous time points throughout both studies. In both studies, the 3.0 RAD group had statistically significantly worse median creatinine and creatinine clearance than those of the MMF group beginning at three months and continuing consistently throughout the remainder of the study at 36 months. Except for creatinine in Study B201 at two time points (30 and 36 months), the 1.5 RAD group had statistically significantly worse median creatinine and creatinine clearance when compared to the MMF group beginning at six months and continuing through 36 months.

1.1 Conclusions and Recommendations

In light of the significant concerns regarding renal toxicities, the inconsistency of the efficacy results (by dose and study), and the possibility that the incidence of graft loss may increase with the use of RAD, it has been suggested that therapeutic drug monitoring may be a more appropriate method for administering RAD. Exploratory analyses of Studies B201 and B251 considering associations between certain achieved drug concentrations and various efficacy and safety parameters could be conducted. However, it is the assessment of this reviewer that such analyses are not sufficient to make confirmatory conclusions as the subjects have not been randomly assigned to their achieved drug concentrations. Therefore, in the assessment of this reviewer, these studies (or re-analyses of them) cannot be used to justify a safe and effective RAD dose or regimen for the prophylaxis of organ rejection in allogeneic kidney transplant patients.

1.2 Brief Overview of Clinical Studies

The sponsor has submitted three controlled Phase 3 studies in this application. Study B253 titled, "A two-year randomized, multicenter, double-blind study of the efficacy and safety of SDZ RAD versus azathioprine as part of a triple immunosuppressive therapy regimen in de novo heart transplant recipients", is intended to support the request for the indication of prophylaxis of organ rejection in heart transplant patients. Two studies, B201 and B251, each titled, "A three-year, double-blind, double dummy, randomized, multicenter, parallel group study of the efficacy and safety of SDZ RAD tablets versus mycophenolate mofetil as part of triple immunosuppressive therapy in de novo renal transplant recipients", were submitted in support of the request for the prophylaxis of organ rejection in allogeneic kidney transplant patients indication. For administrative purposes, this NDA was separated, by indication, into two NDA's thus requiring the assignment of two NDA numbers. The statistical review of the studies intended to support the kidney transplant indication (i.e., studies B201 and B251) is contained in this document. However, the statistical review of the study intended to support the heart transplant indication (i.e., study B253) is completed in a separate document, referenced by NDA 21-628 and 21-631.

The sponsor also submitted synoptic reports of two ongoing studies (A2306 and A2307). Study A2306 was titled, "A 1 year multicenter, randomized, open label, parallel group study of the safety, tolerability and efficacy of two doses (1.5 and 3 mg/day) of Certican (RAD001) with steroids and optimized administration of Neoral in de novo renal transplant recipients". Study A2307 was titled, "A 1 year multicenter, randomized, open label, parallel group study of the safety, tolerability and efficacy of two doses (1.5 and 3 mg/day) of Certican (RAD001) with Simulect, corticosteroids and optimized administration of Neoral in *de novo* renal transplant recipients". Six month synopses were submitted with the 120 Day Safety Update and limited 12 month analyses (describing selected efficacy and safety parameters) were submitted approximately 22 business days prior the User Fee Date for this application. Since these studies did not involve a comparator group (therefore necessitating cross-study comparisons) and since the available information regarding these studies is limited, a description and critique of these studies is not included in this document.

1.3 Statistical Issues and Findings

The following statistical issues are described within the context of the review. Please see the specified references for details.

- Disproportionate premature treatment discontinuation (Ref: Tables 1 and 4, Figures 2 and 4, Sections 3.1.2 and 3.1.3)
- Partial unblinding at six months (Ref: Reviewer's Comment in Section 3.1.3)
- Unblinding and adjustments in study medication at 12 months (Ref: Section 3.1.4)
- Presentation and critique of primary efficacy analyses (Ref: Tables 3 and 4, Figures 3 and 4, Section 3.1.3)

- Presentation of analyses demonstrating a disproportionate deterioration in renal function in the RAD groups (Ref: Table 6 and 7, Figures 5 and 6, Section 3.2)

2. INTRODUCTION

2.1 Overview

The sponsor has submitted the results of two controlled clinical trials (B201 and B251) in support of the efficacy of Certican for prophylaxis of organ rejection in allogeneic kidney transplant patients indication. The studies are each titled, "A three-year, double-blind, double dummy, randomized, multicenter, parallel group study of the efficacy and safety of SDZ RAD tablets versus mycophenolate mofetil as part of triple immunosuppressive therapy in de novo renal transplant recipients". The co-primary objectives of the studies were to demonstrate that the efficacies (as measured by the incidence of biopsy-proven acute allograft rejection, graft loss, death, or lost to follow-up at six months and by the incidence of graft loss, death, or lost to follow-up at 12 months) of at least one of the two oral doses of Certican are non-inferior to those of mycophenolate mofetil (MMF) in *de novo* renal transplant recipients. Since the study design and primary analyses for studies B201 and B251 were nearly identical, an integrated discussion of these trials will be presented. However to allow for interpretation of each study on its own merits, data analyses will be presented separately for each study.

2.2 Data Sources

The sponsor has submitted the results of two controlled clinical trials in support of the efficacy of Certican for prophylaxis of organ rejection in allogeneic kidney transplant patients indication. The following data sets were submitted electronically and utilized in the review of these studies.

\\Cdsesub1\n21560\N_000\2002-12-19\crt\datasets\201\derived\discon.xpt
\\Cdsesub1\n21560\N_000\2002-12-19\crt\datasets\201\derived\efficacy.xpt
\\Cdsesub1\n21560\N_000\2002-12-19\crt\datasets\251\derived\discon.xpt
\\Cdsesub1\n21560\N_000\2002-12-19\crt\datasets\251\derived\efficacy.xpt
\\Cdsesub1\n21560\N_000\2003-08-06\CRT\Datasets\Kidney\derived\crtb201b.xpt
\\Cdsesub1\n21560\N_000\2003-08-06\CRT\Datasets\Kidney\derived\crtb251b.xpt

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

3.1.1 Study Design, Efficacy Endpoints, and Statistical Analysis Plan

B201 and B251 were randomized parallel-group studies intended to compare the efficacy and safety 0.75 mg Certican (RAD) and 1.5 mg Certican (RAD) to 1 gm mycophenolate mofetil (MMF). These double-blind studies were originally designed to last one year; however, were extended (open-label) for two additional years to allow for the collection of data thru 24 and 36 months post-transplant.

Male and female patients, 18 to 68 years of age, who were scheduled to undergo primary cadaveric, living unrelated, or living related (non-HLA identical) kidney transplantation, were to be entered into the studies. Within 48 hours after transplant surgery, subjects were to be randomly assigned to receive one of the following treatments.

Dose Level 1: 0.75 mg RAD bid + Neoral + prednisone
Dose Level 2: 1.5 mg RAD bid + Neoral + prednisone
Comparator: 1 gm MMF bid + Neoral + prednisone

Study B201 was conducted at 48 centers in Europe, four centers in Australia, and two centers in South Africa. Study B251 was conducted at 33 centers in the United States, seven centers in Canada, two centers in Argentina, and two centers in Brazil.

Efficacy assessments were to be collected continuously and in a blinded fashion for three years post-transplantation. However, after all subjects had completed 12 months of therapy and the 12 month analyses of studies B201 and B251 were conducted and reviewed by Novartis, the protocols were amended because of concern regarding renal function. The amendment unblinded the studies and altered the immunosuppressive regimen in an attempt to better manage renal function. Guidelines for altering the regimen as listed in Protocol Amendment #3 are given below.

(1.) Maintain RAD trough levels ≥ 3 ng/mL (so as to minimize the risk of rejection in a maintenance patient), and
(2.) To progressively lower the CsA trough levels to a therapeutic range of 50 – 75 ng / mL (in order to minimize nephrotoxicity).

Prior to entering this phase of the study, subjects were required to sign a revised Informed Consent. Monitoring of subjects was then continued open-label through 36 months post-transplantation.

All efficacy analyses were to be performed using the intent-to-treat (ITT) population, defined as including all patients who were randomized. A selection of the efficacy analyses were to be repeated using a per-protocol (PP) population defined as a subset of the ITT group which satisfied certain protocol specified criteria. Safety and tolerability analyses were to be performed using the safety population defined as all randomized patients who received at least one dose of study medication and then had at least one safety assessment.

The co-primary efficacy endpoints were defined as the proportions of subjects experiencing

(1.) efficacy failure (i.e., biopsy-proven acute rejection episode, graft loss, death, or lost to follow-up) within six months of administration of the initial dose of study medication
(2.) graft loss, death, or lost to follow-up within 12 months of administration of the initial dose of study medication.

The objective of the studies was to demonstrate that *at least one* of the two oral doses of RAD is non-inferior MMF for *both* co-primary efficacy endpoints. For the primary efficacy analysis, the confidence intervals for the by-treatment-group differences in proportions were calculated using normal approximation methods and the non-inferiority criterion, delta, was set a priori at 10%. Adjustments for multiple comparisons (i.e., two doses of RAD) were made using the modified Bonferroni (Hochberg) procedures meaning that if the 95% confidence intervals for the differences in proportions between each RAD group and MMF are within the noninferiority criterion, then both doses of RAD will be considered noninferior to MMF in terms of the co-primary endpoints. Otherwise, each comparison of RAD to MMF will be considered separately using a 97.5% confidence interval for the differences in proportions.

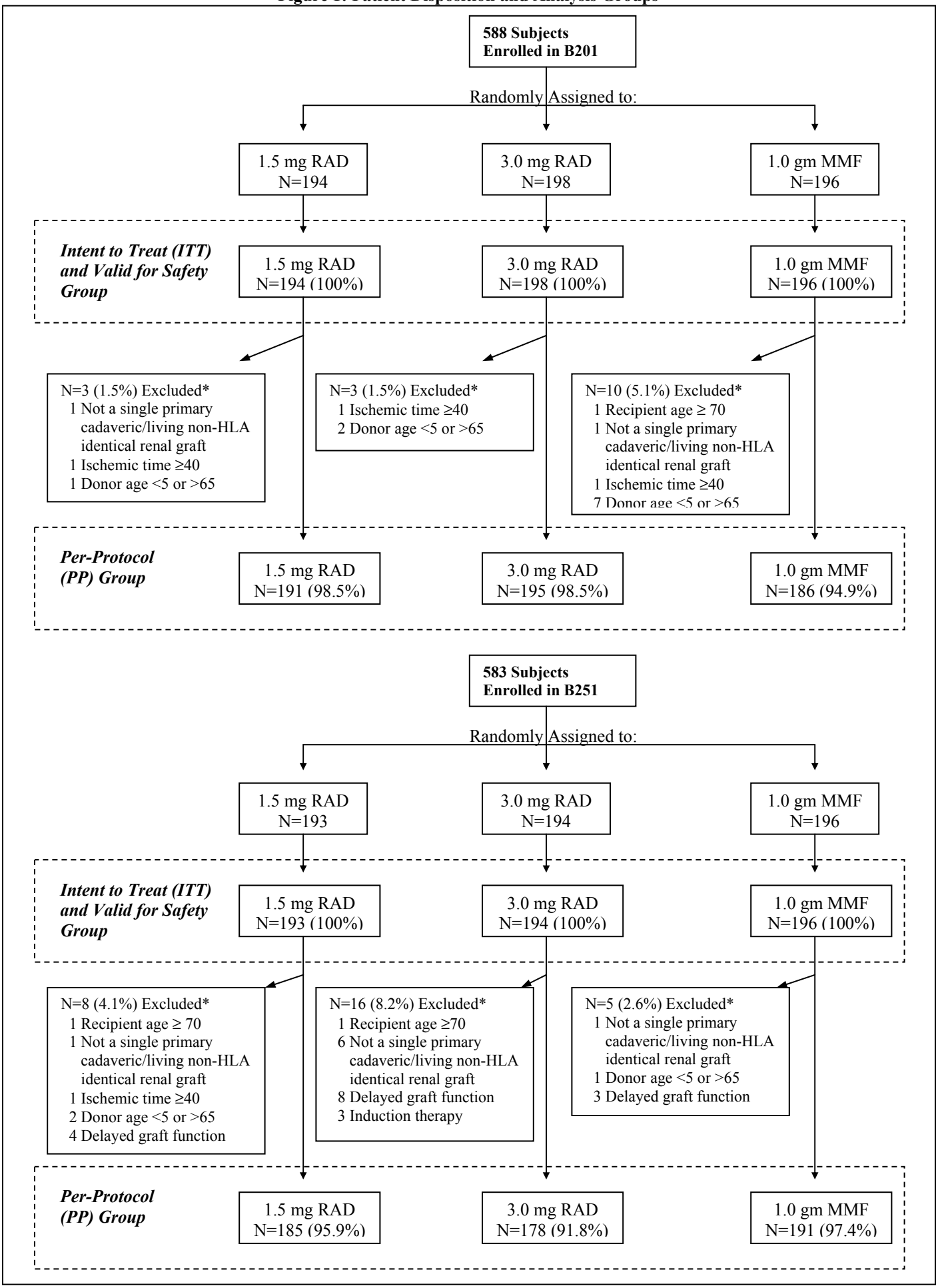
As a supplementary analysis, the Kaplan-Meier estimates of the probability of efficacy failure were compared using the log-rank test to take into account the time to an event.

As specified in the protocol, an interim analysis was performed when 20% of the patients completed three months of the study. The purpose of this analysis was to review the preliminary efficacy data and to investigate any safety problems. The interim analysis was performed by a project-independent internal statistician. Results of this analysis were reviewed by a selected number of decision-making Novartis managers. A significance level of 0.000001 was specified in the protocol and implemented as the alpha level for the interim efficacy analysis at three months. Since the level is set so low, practically, the significance level for the final analysis needs not be adjusted.

3.1.2 Patient Disposition, Demographic and Baseline Characteristics

Five hundred eighty eight subjects and 583 subjects were enrolled in studies B201 and B203, respectively. Figure 1 illustrates the number of subjects randomly assigned to each treatment group and those included in each of the analysis groups. All subjects randomized also received at least one dose of study medication and are therefore included in both the intent-to-treat (ITT) and valid-for-safety analysis groups in both studies B201 and B251. The rate of exclusion of subjects from the per-protocol (PP) analysis group was fairly balanced across treatment groups and was of relatively low frequency. Therefore, the results of the PP analysis group will not be substantially different from those of the ITT analysis group. For this reason and because of the inherent importance of the ITT analysis, this review will focus on the ITT results.

Figure 1: Patient Disposition and Analysis Groups



* Subjects may have more than one reason for exclusion from the PP group.

As indicated in Table 1, approximately 35% of subjects discontinued treatment before Day 450 (i.e., the protocol-defined cutoff date for 12 month safety analyses) in both studies B201 and B251. However, these discontinuations were not balanced across treatment groups. The incidence of premature treatment discontinuation was not statistically significantly different in the two RAD groups; however, the 3.0 mg RAD group had statistically significantly higher premature treatment discontinuation than did the MMF group in study B201 (p=0.1482 for 3.0 RAD versus 1.5 RAD and p=0.0023 for 3.0 RAD versus MMF). The incidence of premature treatment discontinuation was statistically significantly higher in the 3.0 mg RAD group compared with the other two treatment groups in study B251 (p=0.0079 for 3.0 RAD versus 1.5 RAD and p=0.0002 for 3.0 RAD versus MMF). The most common reason for premature treatment discontinuation was adverse events and there were statistically significantly more subjects discontinuing for adverse events in the RAD groups versus the MMF group in B251 (p=0.0207 for 1.5 RAD versus MMF and p=0.0209 for 3.0 RAD versus MMF). The comparison of the rate of discontinuation for adverse events in Study B201 approached statistical significance (p=0.0764) for the 3.0 RAD to MMF comparison. Study discontinuations were infrequent and although numerically more common in the RAD groups, the rates were not statistically significantly different across treatment groups for either study B201 or B251.

Table 1: Premature Treatment or Study Discontinuation (ITT Group)[#]

	Study B201			**Study B251**		
	1.5 RAD (N=194)	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**
Prematurely Discontinued Treatment	**69 (36%)**	**85 (43%)***	**55 (28%)**	**56 (29%)**	**82 (42%)^**	**50 (26%)**
Adverse event(s)	37 (19%)	55 (28%)	39 (20%)	36 (19%)	36 (19%)	20 (10%)
Abnormal lab value(s)	5 (3%)	5 (3%)	1 (1%)	0 (0%)	6 (3%)	3 (2%)
Abnormal test result(s)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)
Unsatisfactory efficacy	17 (9%)	8 (4%)	7 (4%)	13 (7%)	21 (11%)	14 (7%)
Protocol violation	2 (1%)	5 (3%)	4 (2%)	2 (1%)	7 (4%)	4 (2%)
Withdrawal of consent	5 (3%)	9 (5%)	2 (1%)	4 (2%)	8 (4%)	6 (3%)
Lost to follow-up	1 (1%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)
Administrative problems	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)
Death	2 (1%)	3 (2%)	1 (1%)	1 (1%)	3 (2%)	2 (1%)
Prematurely Discontinued Study	**11 (6%)**	**11 (6%)**	**6 (3%)**	**10 (5%)**	**10 (5%)**	**5 (3%)**
Death	10 (5%)	10 (5%)	5 (3%)	6 (3%)	8 (4%)	4 (2%)
Withdrawal of consent	0 (0%)	1 (1%)	1 (1%)	3 (2%)	1 (1%)	0 (0%)
Lost to follow-up	1 (1%)	0 (0%)	0 (0%)	1 (1%)	1 (1%)	1 (1%)

*Statistically significantly differences, p=0.0023 for 3.0 RAD versus MMF

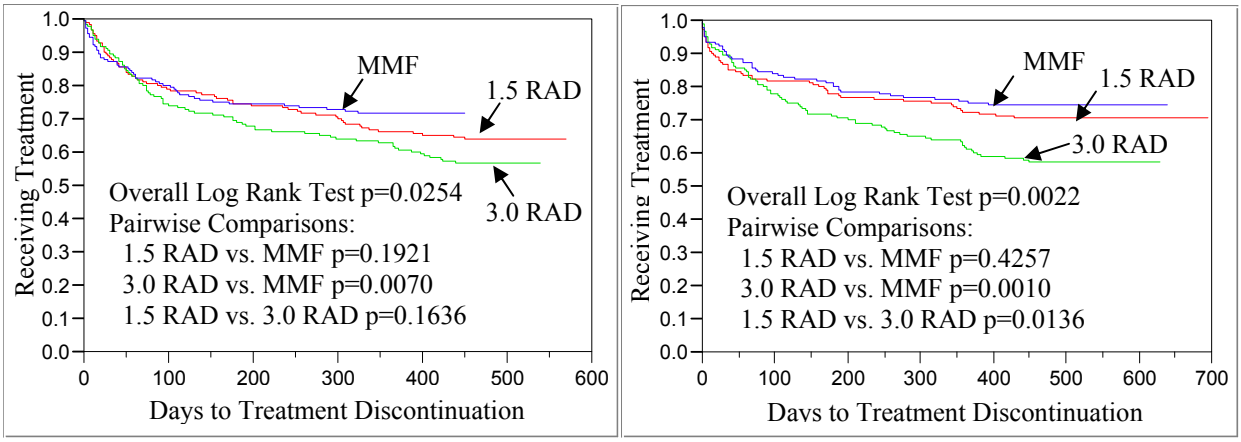
^Statistically significantly differences, p=0.0079 for 3.0 RAD versus 1.5 RAD and p=0.0012 for 3.0 RAD versus MMF

[#]With modifications in format, this table was provided in the sponsor's submission.

A time-to-event analysis of treatment discontinuation is presented in Figure 2 (figure on left corresponds to study B201 and figure on right corresponds to B251). This analysis confirms the previous conclusions indicating that treatment discontinuation occurs statistically significantly earlier and more often in the 3.0 RAD group than it does in the MMF group in both studies.

Figure 2: Time to Treatment Discontinuation by Treatment Group

Study B201 **Study B251**



The disproportionate rate of treatment discontinuation is of concern in the interpretation of both safety and efficacy outcomes in these studies. As more subjects in the 3.0 RAD arm have discontinued treatment the amount of time during which safety events may be reported is by definition shorter (since as per-protocol, adverse events recorded more than 8 days post-treatment discontinuation are not included in the safety analysis). Comparisons of efficacy may be biased by the fact that more subjects in the 3.0 RAD group are receiving alternate therapy than in other treatment groups. After treatment discontinuation, the most commonly used immunosuppressive agents were MMF and Tacrolimus.

The demographic and background characteristics of subjects are summarized in Table 2. In study B201, there were more males than females enrolled in all treatment groups and the majority of patients were Caucasian. The only statistically significant difference observed in this study was the lower proportion of male patients in the 1.5 RAD group as compared to the MMF group. The proportion of cadaveric donors was over 90% in all treatment groups in study B201. In study B251, the majority of patients were Caucasian and less than 50 years of age. The proportion of male patients was significantly lower in the 1.5 RAD group compared with the MMF group, and mean height was significantly lower in the 1.5 RAD group compared with the other treatment groups.

Table 2: Baseline demographics by treatment group (ITT Group)^							
		Study B201			Study B251		
Demographic Variable	**Category / Summary Statistics**	**1.5 RAD (N=194)**	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=194)**	**3.0 RAD (N=198)**	**MMF (N=196)**
Age group, n (%)	**<50 years**	115 (59%)	121 (61%)	112 (57%)	129 (67%)	124 (64%)	126 (64%)
	≥50 years	79 (41%)	77 (39%)	84 (43%)	64 (33%)	70 (36%)	70 (36%)
Age (years)	**Mean (SD)**	45.2 (11.4)	44.1 (11.9)	46.1 (12.3)	43.3 (12.4)	43.7 (12.1)	43.4 (12.1)
	Range	19 to 67	19 to 67	18 to 71	16 to 71	19 to 70	16 to 68
Gender	**Male**	114 (59%)[a]	127 (64%)	139 (71%)	110 (57%)[a]	123 (63%)	132 (67%)
	Female	80 (41%)	71 (36%)	57 (29%)	83 (43%)	71 (37%)	64 (33%)
Race	**Caucasian**	181 (93%)	177 (89%)	171 (87%)	133 (69%)	123 (63%)	129 (66%)
	Black	4 (2%)	9 (5%)	11 (6%)	29 (15%)	36 (19%)	33 (17%)
	Hispanic	0 (0%)	0 (0%)	0 (0%)	20 (10%)	14 (7%)	24 (12%)
	Oriental	4 (2%)	5 (3%)	6 (3%)	3 (2%)	6 (3%)	2 (1%)
	Other	5 (3%)	7 (4%)	8 (4%)	8 (4%)	15 (8%)	8 (4%)
Weight (kg)	**Mean ("SD)**	70.4 (14.3)	70.9 (13.2)	71.2 (13.4)	75.8 (17.5)	76.5 (18.7)	78.7 (17.2)
Height (cm)	**Mean ("SD)**	169.0 (10.4)	170.7 (10.0)	170.8 (9.5)	168.4 (10.6)[a]	170.8 (9.5)[c]	171.0 (10.0)
Diabetes, n (%)	**At Baseline**	8 (4%)	17 (9%)	12 (6%)	36 (19%)	40 (21%)	48 (25%)
Donor source, n (%)	**Cadaveric heart beating**	162 (83%)	169 (85%)	155 (79%)	94 (49%)	93 (48%)	85 (43%)
	Cadaveric non-heart beating	14 (7%)	13 (7%)	23 (12%)	5 (3%)	7 (4%)	5 (3%)
	Living related	9 (5%)	9 (5%)	13 (7%)	62 (32%)	67 (35%)	79 (40%)
	Living unrelated	9 (5%)	7 (4%)	5 (3%)	32 (17%)	27 (14%)	27 (14%)

a:1.5 RAD vs MMF; b: 3.0 RAD vs MMF; and c: 1.5 RAD vs 3.0 RAD ($p<0.05$ using Wilcoxon rank sum test for continuous variables and Fisher's exact test for categorical variables).
^With modifications in format, this table was provided in the sponsor's submission.

3.1.3 Primary Efficacy Results and Conclusions

Recall that the co-primary efficacy endpoints were defined as the proportions of subjects experiencing

(1.) efficacy failure (i.e., biopsy-proven acute rejection episode, graft loss, death, or lost to follow-up) within six months of administration of the initial dose of study medication

(2.) graft loss, death, or lost to follow-up within 12 months of administration of the initial dose of study medication.

Table 3 contains the protocol specified primary analyses of these endpoints.

Table 3: Primary Efficacy Analyses (ITT Group)^						
	Study B201			Study B251		
	1.5 RAD (N=194)	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**
Efficacy Failure Within 6 Months	**52 (26.8%)**	**52 (26.3%)**	**58 (29.6%)**	**42 (21.8%)**	**46 (23.7%)**	**51 (26.0%)**
Biopsy-proven acute rejection	42 (21.6%)	36 (18.2%)	46 (23.5%)	33 (17.1%)	39 (20.1%)	46 (23.5%)
Graft loss	7 (3.6%)	17 (8.6%)	15 (7.7%)	7 (3.6%)	3 (1.5%)	4 (2.0%)
Death	9 (4.6%)	7 (3.5%)	3 (1.5%)	2 (1.0%)	4 (2.1%)	1 (0.5%)
Loss to follow-up	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
95% CI* (RAD – MMF)	**(-11.7%, 6.1%)**[#]	**(-12.2%, 5.6%)**[#]	**NA**	**(-12.7%, 4.3%)**[#]	**(-10.9%, 6.3%)**[#]	**NA**
97.5% CI* (RAD – MMF)	**(-13.0%, 7.4%)**	**(-13.4%, 6.8%)**	**NA**	**(-13.9%, 5.5%)**	**(-12.1%, 7.5%)**	**NA**
Graft Loss, Death, or Loss to Follow-up within 12 Months	**21 (10.8%)**	**33 (16.7%)**	**23 (11.7%)**	**22 (11.4%)**	**15 (7.7%)**	**13 (6.6%)**
Graft loss	9 (4.6%)	21 (10.6%)	18 (9.2%)	17 (8.8%)	8 (4.1%)	10 (5.1%)
Death	10 (5.2%)	8 (4.0%)	5 (2.6%)	6 (3.1%)	7 (3.6%)	4 (2.0%)
Lost to follow-up	3 (1.5%)	4 (2.0%)	2 (1.0%)	1 (0.5%)	1 (0.5%)	1 (0.5%)
95% CI* (RAD – MMF)	**(-7.2%, 5.4%)**	**(-1.9%, 11.9%)**	**NA**	**(-0.9%, 10.5%)**	**(-4.0%, 6.2%)**	**NA**
97.5% CI* (RAD – MMF)	**(-8.1%, 6.3%)**[#]	**(-2.9%, 12.9%)**	**NA**	**(-1.7%, 11.3%)**	**(-4.7%, 6.9%)**[#]	**NA**

*Calculated using normal approximation methods.
^Individual components of the efficacy endpoint are all presented and may not be mutually exclusive.
[#]Shaded areas indicate statistically noninferior results.

Using the protocol defined delta of 10% and the modified Bonferroni (Hochberg) multiple comparison procedure (previously described in Section 3.1.1), the results in Table 3 indicate that for Study 201, the rate of efficacy failure within six months for each of the RAD groups are noninferior to that of MMF (as evidenced by the fact that the upper limits of both 95% confidence intervals are within 10%). Also in Study 201, but for the graft loss, death, or loss to follow-up endpoint at 12 months, only the 1.5 RAD group is shown to be noninferior to the MMF group (as evidenced by the fact that the upper limit of the 95% confidence interval for 3.0 RAD minus MMF does not exclude 10% while the 97.5% confidence interval for the 1.5 RAD minus MMF does exclude 10%).

For Study 251, the rates of efficacy failure within six months for each of the RAD groups are shown to be noninferior to that of MMF (as evidenced by the fact that the upper limits of both 95% confidence intervals are within 10%). Also in Study 251, but for the graft loss, death, or loss to follow-up endpoint at 12 months, only the 3.0 RAD group is shown to be noninferior to the MMF group (as evidenced by the fact that the upper limit of the 97.5% confidence interval for 3.0 RAD minus MMF does exclude 10%).

It should be noted that at the time the six month analysis was conducted, Novartis statisticians and programmers were unblinded to all data, and selected Novartis clinical research personnel were unblinded to the results of the six month analyses. However, the sponsor's study report states that the patients, investigators, and study center personnel did remain blinded to the data.

Reviewer's Comment: While the study report indicates that patients, investigators, and study center personnel remained blinded to the efficacy analyses conducted at six months, revealing efficacy results such as these to persons who are even peripherally involved in the study is not appropriate as it may have biased the results obtained beyond six months. If it was felt that ethically the six month results needed to be reviewed to allow the progression of a blinded trial, establishment a Data Safety Monitoring Board for this purpose would have been more appropriate. It is not possible to quantify the magnitude of bias that may have been caused as a result of this unblinding.

While the primary efficacy results in Table 3 appear somewhat promising, one should remain mindful of the fact that there were disproportionate rates of treatment discontinuation and that comparisons of efficacy may be biased by the fact that more subjects in the 3.0 RAD group were receiving alternate therapy than in other treatment groups. In order to address this issue an analysis considering treatment discontinuation failures of therapy in conjunction with the previously defined co-primary efficacy endpoints is presented in Table 4.

Table 4: Primary Efficacy Endpoints with Premature Treatment Discontinuation Considered a Failure (ITT Group)						
	Study B201			Study B251		
	1.5 RAD (N=194)	3.0 RAD (N=198)	MMF (N=196)	1.5 RAD (N=193)	3.0 RAD (N=194)	MMF (N=196)
Efficacy Failure or Premature Treatment Discontinuation Within Six Months	71 (36.6%)	79 (39.9%)	74 (37.8%)	62 (32.1%)	70 (36.1%)	66 (33.7%)
95% CI* (RAD – MMF)	(-10.7%, 8.4%)	(-7.5%, 11.7%)	NA	(-10.9%, 7.8%)	(-7.0%, 11.8%)	NA
97.5% CI* (RAD – MMF)	(-12.1%, 9.8%)#	(-8.8%, 13.1%)	NA	(-12.2%, 9.1%)#	(-8.4%, 13.2%)	NA
Graft Loss, Death, Loss to Follow-up, or Premature Treatment Discontinuation Within 12 Months	65 (33.5%)	77 (38.9%)	55 (28.1%)	54 (28.0%)	78 (40.2%)	48 (24.5%)
95% CI* (RAD – MMF)	(-3.7%, 14.6%)	(1.5%, 20.0%)	NA	(-5.3%, 12.2%)	(6.5%, 24.8%)	NA
97.5% CI* (RAD – MMF)	(-5.0%, 15.8%)	(0.2%, 21.3%)	NA	(-6.5%, 13.5%)	(5.1%, 26.0%)	NA

*Calculated using normal approximation methods.
#Shaded areas indicate statistically noninferior results.

The rate of first occurrence of efficacy failure or treatment discontinuation within six months for the 1.5 RAD group is noninferior to that of MMF in both Studies 201 and 251 (as evidenced by the fact that the upper limits of the 97.5% confidence intervals for 1.5 RAD minus MMF exclude 10%). However, this is not the case for the 3.0 RAD group. Noninferiority of the rates of first occurrence of efficacy failure or treatment discontinuation at six months for the 3.0 RAD and MMF groups has not been demonstrated in either study.

Partially because the rates of graft loss, death, and lost-to-follow up at 12 months were low relative to those of premature treatment discontinuation and since larger proportions of RAD subjects discontinued treatment, neither of the RAD groups were shown to be noninferior to MMF for this endpoint in either study. In fact, the rates of first occurrence of graft loss, death, lost-to-follow-up, or treatment discontinuation in the 3.0 RAD group were shown to be statistically significantly higher than those of the MMF group in both studies.

Figures 3 and 4 provide a visual display of the data in Tables 3 and 4. The Kaplan Meier plots for co-primary efficacy endpoints, efficacy failure at six months and death, graft loss, or lost-to-follow-up at 12 months, are given in Figure 3. Using time-to-event analyses (i.e., log rank test), no statistically significant differences were found for these endpoints in either Studies 201 or 251. Figure 4 contains the Kaplan Meier plots for the co-primary efficacy endpoints with premature treatment discontinuation considered an event (i.e., as a failure of therapy). With one exception (i.e., 3.0 RAD group compared to MMF for the 12 month endpoint in study 251) no statistically significant differences were found in these analyses in either study. Visual comparisons from Figure 3 to Figure 4 illustrate the effect of considering premature treatment discontinuation a failure of therapy.

Figure 3: Kaplan-Meier Estimates for Co-Primary Efficacy Endpoints by Treatment Group

Study B201 **Study B251**

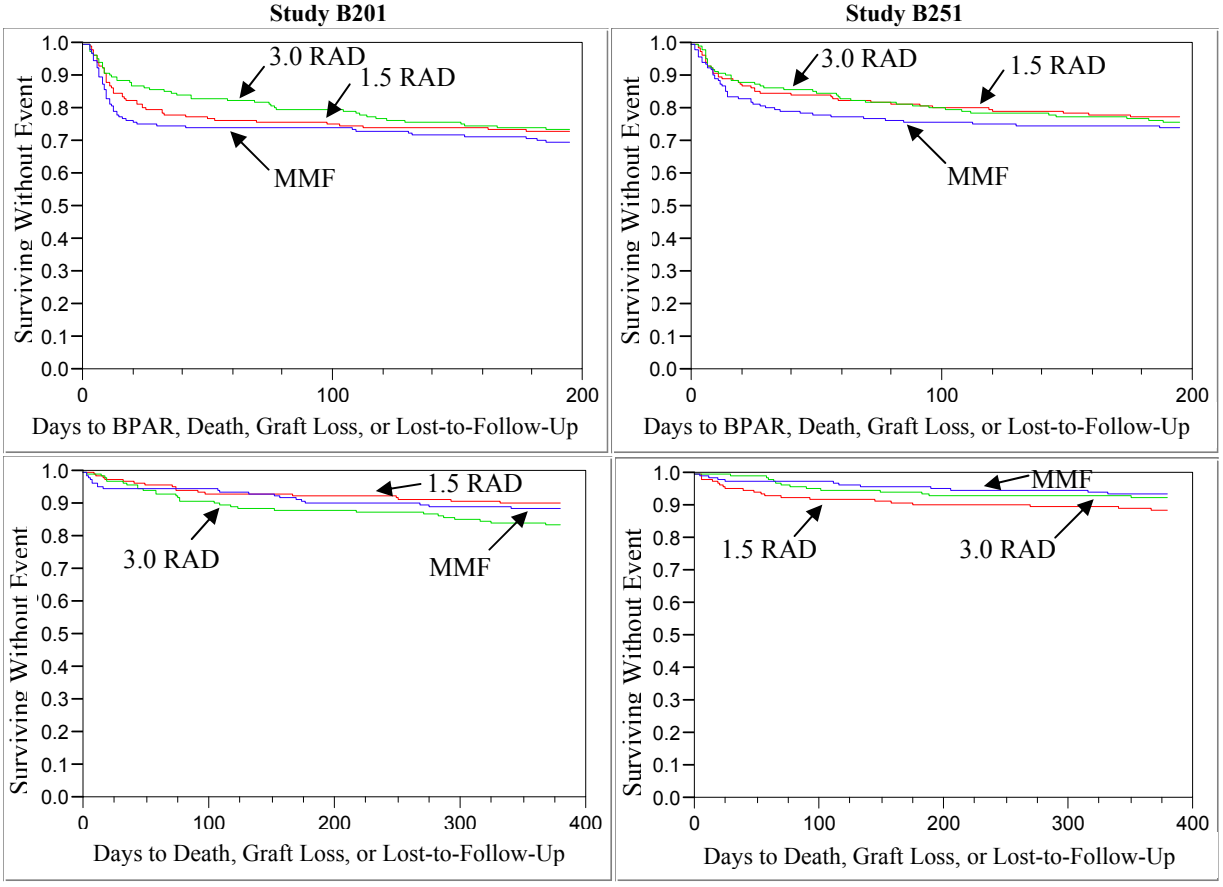
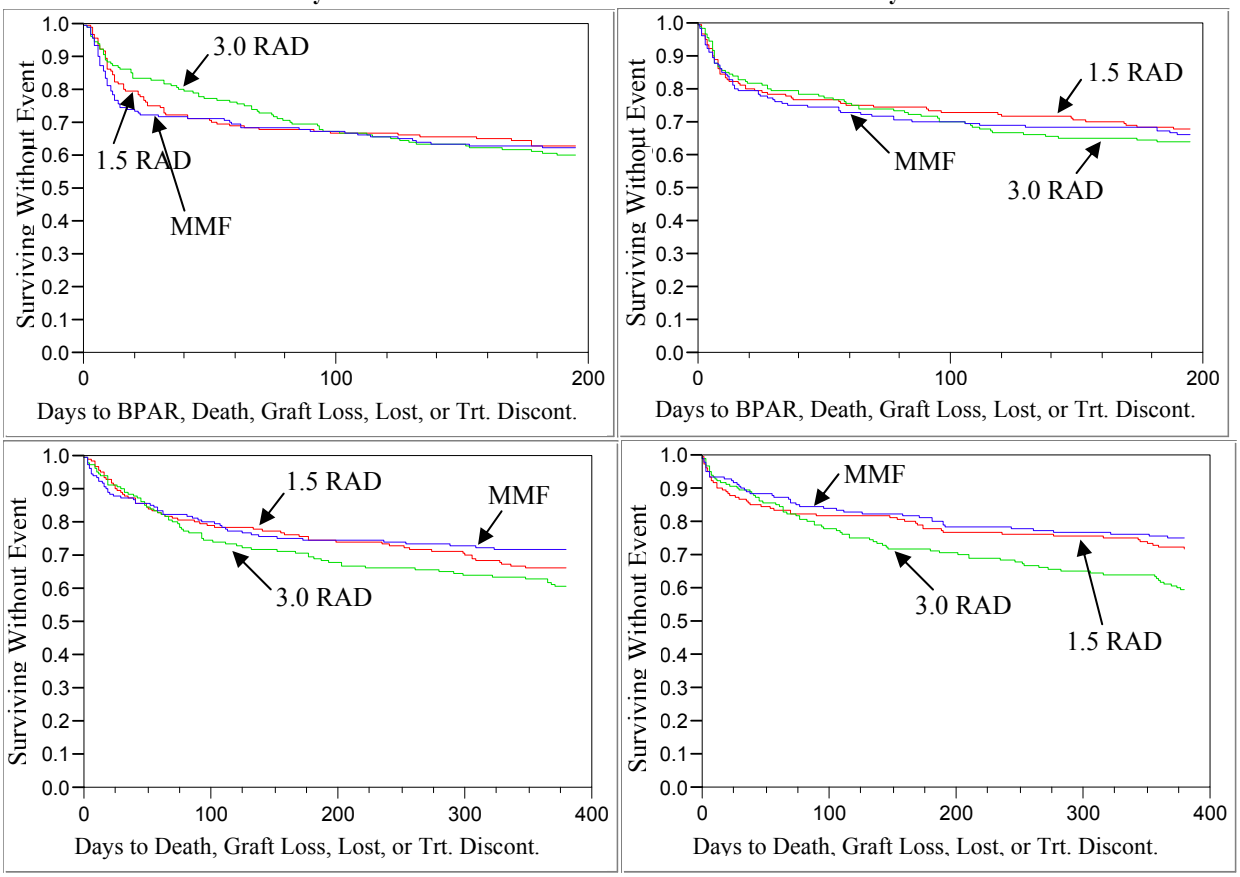


Figure 4: Kaplan-Meier Estimates for Co-Primary Efficacy Endpoints or Treatment Discontinuation by Trt. Group

Study B201 **Study B251**



3.1.4 Secondary Efficacy Results and Conclusions

This section is dedicated to the examination of selected secondary efficacy results including by-treatment comparisons of each component of the primary efficacy endpoints as well as analyses at time points other than those defined as primary.

Recall that these double-blind studies were originally designed to last one year but were extended (open-label) for two additional years. Efficacy assessments were to be collected continuously and in a blinded fashion for three years post-transplantation. However, after all subjects had completed 12 months of therapy and the 12 month analyses of studies B201 and B251 were conducted and reviewed by Novartis, the protocols were amended because of concern regarding renal function. The amendment unblinded the studies and altered the immunosuppressive regimen in an attempt to better manage renal function. Monitoring of subjects was then continued open-label through 36 months post-transplantation. Because of the unblinding and/or dose adjustments at 12 months, subjective efficacy results beyond the 12 month time point are considered by this reviewer to be more subject to bias and less reliable than those prior to this time point. However, this bias may have little impact on objective endpoints, such as graft loss or death. These results are provided in Table 5A and 5B as secondary efficacy analyses.

The results in Table 5A for Study B201 indicate that the rates of the co-primary composites (biopsy-proven acute rejection, graft loss, death, or loss-to-follow-up; and graft loss, death, or loss-to-follow-up) for the 1.5 RAD group are non-inferior to those of the MMF group for all time points listed. In addition, examination of the composites' components at each time point consistently indicates a non-inferior result for 1.5 RAD compared to MMF. However, the same pattern is not as clear for the 3.0 RAD to MMF comparison. By 12 months, the non-inferiority of the rate of graft loss, death, or lost-to-follow-up in the 3.0 RAD group to that of the MMF group is not maintained. In addition, at 36 months, the 3.0 RAD group has a statistically significantly higher rate of graft loss, death, or loss-to-follow-up when compared to the MMF group (25.3% versus 16.3%, respectively). Considering each of the components of this composite alone, it is evident that the event of graft loss is the most problematic in this regard. At 36 months, despite the overall low incidence, the rate of graft loss in the 3.0 RAD group is nearly statistically significantly higher than that of the MMF group (16.7% versus 10.7%, respectively).

Unfortunately, the dose-response relationships observed in Study B201 are not replicated in Study B251. In this study, the comparison of the rate of graft loss, death, or loss-to-follow-up for the 1.5 RAD group to that of the MMF group did not meet the non-inferiority criteria for any of the time points. And the non-inferiority of 1.5 RAD to MMF in terms of the biopsy proven acute rejection, death, graft loss, or lost-to-follow-up composite is not maintained through the 36 month time point. Examination of each of the components of the composites indicate that the event of graft loss is the most problematic in this regard and at 36 months, the rate of graft loss in the 1.5 RAD group is not non-inferior to that of the MMF group. While comparison of the 3.0 RAD group to the MMF group in terms of the co-primary composites yields non-inferior results at the 6 and 12 month time points, this non-inferiority is not maintained through the 36 month time point for either co-primary endpoint.

Table 5A: Secondary Efficacy Analyses (ITT Group) for Study B201

	6 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)			12 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)			36 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)		
Endpoint	**1.5 RAD (N=194)**	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=194)**	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=194)**	**3.0 RAD (N=198)**	**MMF (N=196)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**52 (26.8%) (-11.7%, 6.1%) (-13.0%, 7.4%)**	**52 (26.3%) (-12.2%, 5.6%) (-13.4%, 6.8%)**	**58 (29.6%) NA NA**	**58 (29.9%) (-10.3%, 7.9%) (-11.7%, 9.3%)**	**60 (30.3%) (-9.9%, 8.3%) (-11.2%, 9.6%)**	**61 (31.1%) NA NA**	**64 (33.0%) (-13.7%, 5.2%) (-15.0%, 6.6%)**	**77 (38.9%) (-7.9%, 11.2%) (-9.3%, 12.5%)**	**73 (37.2%) NA NA**
Graft Loss, Death, or Loss to Follow-up Composite	**15 (7.7%) (-7.8%, 3.8%) (-8.7%, 4.7%)**	**24 (12.1%) (-3.8%, 8.8%) (-4.8%, 9.7%)**	**19 (9.7%) NA NA**	**21 (10.8%) (-7.2%, 5.4%) (-8.1%, 6.3%)**	**33 (16.7%) (-1.9%, 11.9%) (-2.9%, 12.9%)**	**23 (11.7%) NA NA**	**27 (13.9%) (-9.5%, 4.7%) (-10.5%, 5.7%)**	**50 (25.3%) (1.0%, 17.0%) (-0.1%, 18.1%)**	**32 (16.3%) NA NA**
Biopsy-Proven Acute Rejection (single event)	42 (21.6%) (-10.1%, 6.5%) (-11.3%, 7.7%)	36 (18.2%) (-13.3%, 2.8%) (-14.5%, 3.9%)	46 (23.5%) NA NA	45 (23.2%) (-9.2%, 7.6%) (-10.4%, 8.8%)	39 (19.7%) (-12.5%, 3.9%) (-13.6%, 5.0%)	47 (24.0%) NA NA	47 (24.2%) (-10.9%, 6.3%) (-12.2%), 7.6%)	49 (24.7%) (-10.4%, 6.8%) (-11.7%, 8.1%)	52 (26.5%) NA NA
Graft loss (single event)	7 (3.6%) (-9.1%, 0.6%) (-9.9%, 1.3%)	17 (8.6%) (-4.7%, 6.6%) (-5.5%, 7.4%)	15 (7.7%) NA NA	9 (4.6%) (-9.6%, 0.4%) (-10.3%, 1.1%)	21 (10.6%) (-4.5%, 7.3%) (-5.3%, 8.1%)	18 (9.2%) NA NA	14 (7.2%) (-9.2%, 2.2%) (-10.0%, 3.0%)	33 (16.7%) (-0.8%, 12.8%) (-1.7%, 13.7%)	21 (10.7%) NA NA
Death (single event)	9 (4.6%) (-0.4%, 7.2%) (-1.0%, 8.0%)	7 (3.5%) (-1.3%, 5.8%) (-1.9%, 6.5%)	3 (1.5%) NA NA	10 (5.2%) (-1.2%, 6.4%) (-1.8%, 7.0%)	8 (4.0%) (-2.1%, 4.9%) (-2.6%, 5.4%)	5 (2.6%) NA NA	15 (7.7%) (-5.9%, 4.9%) (-6.6%, 5.6%)	18 (9.1%) (-4.6%, 6.4%) (-5.4%, 7.2%)	16 (8.2%) NA NA

Table 5B: Secondary Efficacy Analyses (ITT Group) for Study B251

	6 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)			12 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)			36 Months freq (%), 95% CI*, 97.5% CI* (RAD-MMF)		
Endpoint	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**
Biopsy-Proven Acute Rejection, Graft Loss, Death, or Loss to Follow-up Composite	**42 (21.8%) (-12.7%, 4.3%) (-13.9%, 5.5%)**	**46 (23.7%) (-10.9%, 6.3%) (-12.1%, 7.5%)**	**51 (26.0%) NA NA**	**48 (24.9%) (-11.4%, 6.0%) (-12.7%, 7.3%)**	**51 (26.3%) (-10.1%, 7.5%) (-11.4%, 8.8%)**	**54 (27.6%) NA NA**	**65 (33.7%) (-6.7%, 11.8%) (-8.1%, 13.1%)**	**66 (34.0%) (-6.4%, 12.2%) (-7.7%, 13.5%)**	**61 (31.1%) NA NA**
Graft Loss, Death, or Loss to Follow-up Composite	**19 (9.8%) (-0.5%, 10.4%) (1.3%, 11.2%)**	**13 (6.7%) (-3.3%, 6.6%) (-4.1%, 7.5%)**	**10 (5.1%) NA NA**	**22 (11.4%) (-0.9%, 10.5%) (-1.7%, 11.3%)**	**15 (7.7%) (-4.0%, 6.2%) (-4.7%, 6.9%)**	**13 (6.6%) NA NA**	**33 (17.1%) (-1.0%, 12.8%) (-2.0%, 13.8%)**	**28 (14.4%) (-3.4%, 9.8%) (-4.4%, 10.8%)**	**22 (11.2%) NA NA**
Biopsy-Proven Acute Rejection (single event)	33 (17.1%) (-14.4%, 1.6%) (-15.5%, 2.8%)	39 (20.1%) (-11.6%, 4.9%) (-12.8%, 6.1%)	46 (23.5%) NA NA	37 (19.2%) (-13.0%, 3.4%) (-14.1%, 4.5%)	43 (22.2%) (-10.2%, 6.6%) (-11.4%, 7.8%)	47 (24.0%) NA NA	49 (25.4%) (-9.8%, 7.6%) (-11.1%, 8.9%)	50 (25.8%) (-9.4%, 8.0%) (-10.7%, 9.3%)	52 (26.5%) NA NA
Graft loss (single event)	15 (7.8%) (-0.4%, 9.3%) (-1.2%, 10.1%)	7 (3.6%) (-4.0%, 4.1%) (-4.8%, 4.9%)	7 (3.6%) NA NA	17 (8.8%) (-1.3%, 8.7%) (-2.1%, 9.5%)	8 (4.1%) (-5.2%, 3.2%) (-5.8%, 3.8%)	10 (5.1%) NA NA	23 (11.9%) (-1.0%, 10.6%) (-1.8%, 11.4%)	15 (7.7%) (-4.6%, 5.8%) (-5.3%, 6.5%)	14 (7.1%) NA NA
Death (single event)	5 (2.6%) (-1.4%, 5.0%) (-2.0%, 5.7%)	6 (3.1%) (-0.9%, 5.7%) (-1.5%, 6.4%)	2 (1.0%) NA NA	6 (3.1%) (-2.0%, 4.2%) (-2.5%, 4.7%)	7 (3.6%) (-1.7%, 4.9%) (-2.1%, 5.3%)	4 (2.0%) NA NA	12 (6.2%) (-3.5%, 5.7%) (-4.1%, 6.3%)	13 (6.7%) (-3.1%, 6.3%) (-3.7, 6.9%)	10 (5.1%) NA NA

*Calculated using normal approximation methods.

3.2 Evaluation of Safety

Because of concerns regarding possible renal toxicities associated with 1.5 RAD and/or 3.0 RAD, this section is primarily devoted to the exploration of the creatinine and creatinine clearance endpoints. However, it should first be noted that as previously described in Section 3.1.2 of this document there were more subjects prematurely discontinuing study treatment due to adverse events in the RAD groups versus the MMF group. In Study B201 these differences did not reach statistical significance. But in Study B251, the rates of premature treatment discontinuation in the 1.5 RAD and 3.0 RAD groups were statistically significantly higher than that of the MMF group (18.7%, 18.6%, and 10.2%, respectively, 1.5 RAD vs. MMF p=0.0207, 3.0 RAD vs. MMF p=0.0209). Therefore there were disproportionately more subjects in the RAD groups receiving alternate therapy than in the MMF group. Since the renal function analyses presented herein include subjects who are no longer receiving their randomly assigned treatment, this may have lead to a disproportionate bias and should be kept in mind in the interpretation of the creatinine and creatinine clearance endpoints. For a comprehensive discussion of the safety of RAD, please refer to the Medical Review and Evaluation for NDA 21-560.

Median creatinine and creatinine clearance (calculated using the Nankivell method) and by-treatment group comparisons of these endpoints are presented in Table 6 and 7, respectively. Statistically significant differences among treatment groups were obsevered at many time points and are highlighted in the tables by bold-facing the p-value for the comparison. In both studies, the 3.0 RAD group had statistically significantly worse median creatinine and creatinine clearance than those of the MMF group beginning at three months and continuing consistently throughout the remainder of the study at 36 months. Changes in the dosage regimen at 12 months did not appear to stabilize the differences in renal function between the 3.0 RAD and MMF groups. Except for creatinine in Study B201 at time points 30 months and 36 months, the 1.5 RAD group had statistically significantly worse median creatinine and creatinine clearance when compared to the MMF group beginning at six months and continuing through 36 months. Changes in the dosage regimen at 12 months did not appear to stabilize the differences in renal function between the 1.5 RAD and MMF groups. Graphical representations across time of the median creatinine and creatinine clearance support the conclusions already drawn from Table 6 and 7 and are included in Figures 5 and 6, respectively.

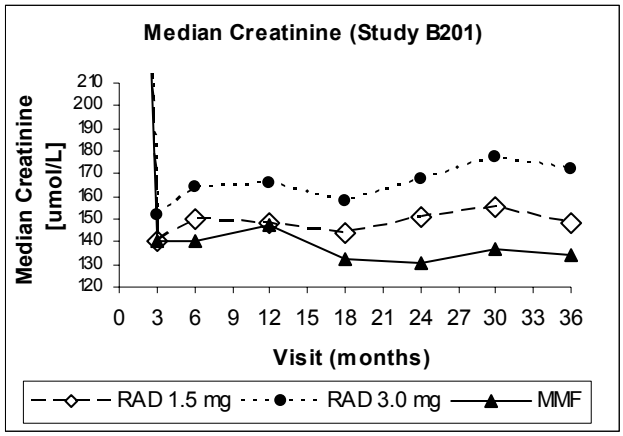
Table 6: Median Creatinine (µmol/L) by Treatment Group (ITT Group*)						
	Study B201			Study B251		
	1.5 RAD (N=194)	3.0 RAD (N=198)	MMF (N=196)	1.5 RAD (N=193)	3.0 RAD (N=194)	MMF (N=196)
Baseline	633 (N=189)	655 (n=190)	656 (n=180)	501 (N=190)	480 (N=189)	453 (N=189)
RAD vs. MMF^	p=0.323	p=0.619	NA	p=0.460	p=0.664	NA
1.5 RAD vs. 3.0 RAD^	p=0.596	NA	NA	p=0.776	NA	NA
Month 3	140 (N=154)	152 (N=152)	140 (N=158)	133 (N=155)	151 (N=152)	133 (N=159)
RAD vs. MMF^	p=0.737	**p=0.004**	NA	p=0.472	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	**p=0.014**	NA	NA	**p=0.001**	NA	NA
Month 6	150 (N=146)	164 (N=135)	140 (N=147)	142 (N=150)	159 (N=137)	133 (N=152)
RAD vs. MMF^	**p=0.034**	**p<0.001**	NA	**p=0.020**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	**p=0.017**	NA	NA	**p=0.012**	NA	NA
Month 12	148 (N=124)	166 (N=120)	147 (N=138)	142 (N=141)	160 (N=117)	133 (N=142)
RAD vs. MMF^	**p=0.037**	**p<0.001**	NA	**p=0.003**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.063	NA	NA	**p=0.003**	NA	NA
Month 18	144 (N=117)	158 (N=100)	132 (N=132)	142 (N=125)	169 (N=100)	128 (N=130)
RAD vs. MMF^	**p=0.010**	**p<0.001**	NA	**p=0.003**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.064	NA	NA	**p=0.001**	NA	NA
Month 24	151 (N=100)	168 (N=89)	131 (N=118)	148 (N=114)	204 (N=85)	128 (N=121)
RAD vs. MMF^	**p=0.009**	**p<0.001**	NA	**p=0.011**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	**p=0.035**	NA	NA	**p<0.001**	NA	NA
Month 30	155 (N=92)	177 (N=87)	137 (N=111)	151 (N=105)	195 (N=78)	133 (N=108)
RAD vs. MMF^	p=0.057	**p<0.001**	NA	**p=0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	**p=0.004**	NA	NA	**p<0.001**	NA	NA
Month 36	148 (N=92)	172 (N=83)	134 (N=108)	145 (N=87)	174 (N=65)	124 (N=108)
RAD vs. MMF^	p=0.069	**p<0.001**	NA	**p=0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	**p=0.016**	NA	NA	**p=0.005**	NA	NA

*Measurements for subjects who are no longer receiving randomly assigned treatment are included in this analysis.
^Wilcoxon Rank-Sum test

Table 7: Median Estimated Creatinine Clearance (mL/min) using Nankivell Method by Treatment Group (ITT Group*)						
	Study B201			Study B251		
	1.5 RAD (N=194)	3.0 RAD (N=198)	MMF (N=196)	1.5 RAD (N=193)	3.0 RAD (N=194)	MMF (N=196)
Baseline	18.5 (N=184)	18.7 (N=188)	18.6 (N=178)	23.7 (N=187)	24.3 (N=185)	26.8 (N=184)
RAD vs. MMF^	p=0.627	p=0.887	NA	**p=0.039**	p=0.116	NA
1.5 RAD vs. 3.0 RAD^	p=0.727	NA	NA	p=0.653	NA	NA
Month 3	57.3 (N=154)	54.9 (N=152)	60.0 (N=158)	61.8 (N=155)	58.1 (N=151)	64.0 (N=159)
RAD vs. MMF^	p=0.168	**p=0.004**	NA	p=0.060	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.119	NA	NA	p=0.057	NA	NA
Month 6	56.7 (N=146)	52.9 (N=135)	61.0 (N=147)	58.4 (N=150)	54.9 (N=135)	65.6 (N=151)
RAD vs. MMF^	**p=0.003**	**p<0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.197	NA	NA	p=0.195	NA	NA
Month 12	54.3 (N=123)	53.3 (N=119)	60.3 (N=138)	58.0 (N=140)	55.2 (N=116)	66.6 (N=141)
RAD vs. MMF^	**p=0.002**	**p<0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.389	NA	NA	p=0.247	NA	NA
Month 18	56.2 (N=117)	54.3 (N=100)	64.2 (N=132)	59.8 (N=124)	53.0 (N=96)	68.0 (N=128)
RAD vs. MMF^	**p<0.001**	**p<0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.467	NA	NA	**p=0.031**	NA	NA
Month 24	57.8 (N=100)	53.0 (N=88)	63.5 (N=118)	62.0 (N=113)	51.9 (N=83)	70.4 (N=119)
RAD vs. MMF^	**p=0.005**	**p<0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.082	NA	NA	**p=0.010**	NA	NA
Month 30	56.7 (N=92)	55.7 (N=87)	64.1 (N=110)	60.2 (N=103)	54.9 (N=77)	71.2 (N=108)
RAD vs. MMF^	**p=0.005**	**p<0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.148	NA	NA	**p=0.033**	NA	NA
Month 36	57.8 (N=89)	56.3 (N=79)	65.1 (N=101)	58.4 (N=86)	53.2 (N=63)	71.7 (N=104)
RAD vs. MMF^	**p=0.007**	**p=0.001**	NA	**p<0.001**	**p<0.001**	NA
1.5 RAD vs. 3.0 RAD^	p=0.425	NA	NA	P=0.357	NA	NA

*Measurements for subjects who are no longer receiving randomly assigned treatment are included in this analysis.
^Wilcoxon Rank-Sum test

Figure 5



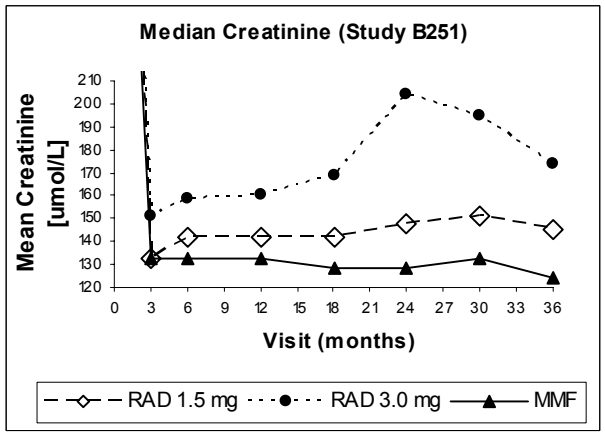
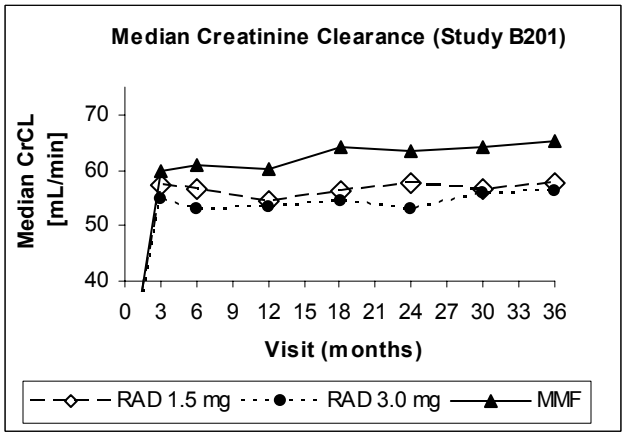
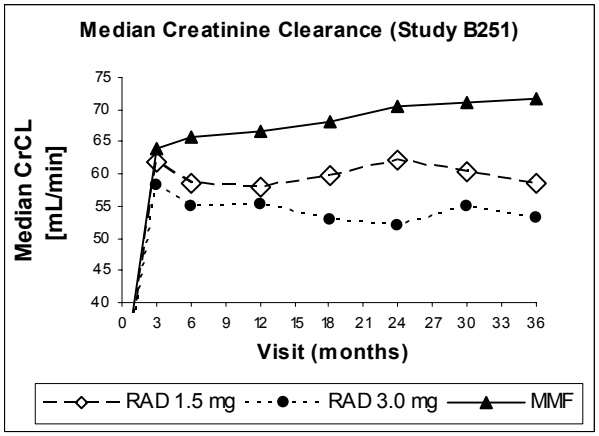


Figure 6





At the request of the medical review team, an additional analysis of creatinine clearance was conducted and is included in Table 8. For this analysis, a subject's "baseline" creatinine clearance is defined as the creatinine clearance value corresponding to the lowest on-treatment creatinine value within the first month after transplant. This value is expected to represent an individual's best post-transplantation creatinine clearance. The analyses in Table 8 compare the proportions of subjects with creatinine clearance values falling below 50% of their baseline creatinine clearance value. A threshold of 50% was chosen by the medical team as being indicative of postoperative acute renal failure.

Before interpretation of the results in Table 8, it should be noted that because a subject's "baseline" value was achieved while on study treatment, the effect of study treatment in altering that subject's "baseline" value cannot be ruled out. In fact, as illustrated in Figure 6, it is likely that 1.5 and 3.0 RAD subjects had lower creatinine clearance values at "baseline" than did the MMF subjects. Dividing by an artificially small "baseline" value (to create the endpoint of interest) would result in an artificially large proportion. In other words, a more extreme result post-baseline would be needed in order for the RAD groups to fall below 50% of "baseline" than would be needed in the MMF group. Nonetheless, in Study B201, the proportion of subjects with creatinine clearance values

less than 50% of their "baseline" value is statistically significantly higher in the 1.5 RAD group when compared to the MMF group at nine, 12, and 18 months (indicated by shaded areas). Also in Study B201, the proportion of subjects with creatinine clearance values less than 50% of their "baseline" value is statistically significantly higher in the 3.0 RAD group when compared to the MMF group at nine, 12, 18, 24, and 36 months (indicated by shaded areas). For Study B251, statistically significant by-treatment group comparisons were observed only sporadically. As previously discussed, this lack of statistical significance may be due in part to the artificially low "baseline" values in the RAD groups.

Table 8: Subjects with Creatinine Clearance ≤ 50% of "Baseline"*

	Study B201			Study B251		
	1.5 RAD (N=194)	**3.0 RAD (N=198)**	**MMF (N=196)**	**1.5 RAD (N=193)**	**3.0 RAD (N=194)**	**MMF (N=196)**
Month 3	5/180 (2.8%)	8/177 (4.5%)	2/184 (1.1%)	4/174 (2.3%)	1/180 (0.6%)	3/183 (1.6%)
RAD vs. MMF^	p=0.2797	p=0.0575	NA	p=0.7177	p=0.6230	NA
Month 6	7/177 (4.0%)	4/172 (2.3%)	2/186 (1.1%)	3/167 (1.8%)	2/174 (1.1%)	5/178 (2.8%)
RAD vs. MMF^	p=0.0976	p=0.4331	NA	p=0.7245	p=0.4486	NA
Month 9	7/141 (5.0%)	10/128 (7.8%)	0/159 (0.0%)	6/142 (4.2%)	4/124 (3.2%	2/140 (1.4%)
RAD vs. MMF^	p=0.0047	p=0.0003	NA	p=0.2821	p=0.4242	NA
Month 12	9/168 (5.4%)	10/167 (6.0%)	1/180 (0.6%)	6/173 (3.5%)	5/166 (3.0%)	3/173 (1.7%)
RAD vs. MMF^	p=0.0086	p=0.0043	NA	p=0.5020	p=0.4946	NA
Month 18	4/119 (3.4%)	6/106 (5.7%)	0/135 (0.0%)	7/128 (5.5%)	8/105 (7.6%)	2/134 (1.5%)
RAD vs. MMF^	p=0.0469	p=0.0067	NA	p=0.0967	p=0.0237	NA
Month 24	4/160 (2.5%)	11/151 (7.3%)	3/161 (1.9%)	6/152 (3.9%)	7/140 (5.0%)	2/155 (1.3%)
RAD vs. MMF^	p=0.7231	p=0.0272	NA	p=0.1705	p=0.0904	NA
Month 30	2/99 (2.0%)	7/92 (7.6%)	2/113 (1.8%)	7/105 (6.7%)	4/84 (4.8%)	3/114 (2.6%)
RAD vs. MMF^	p=1.0000	p=0.0813	NA	p=0.2007	p=0.4603	NA
Month 36	3/158 (1.9%)	8/148 (5.4%)	1/159 (0.6%)	9/141 (6.4%)	7/122 (5.7%)	2/148 (1.4%)
RAD vs. MMF^	p=0.3710	p=0.0163	NA	p=0.0315	p=0.0834	NA

*"Baseline" creatinine clearance is defined as the creatinine clearance value corresponding to the lowest on-treatment creatinine value within the first month after transplant.
^Fisher's Exact Test.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

Subgroup analyses of efficacy failure (i.e., biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) by gender, age, and race were conducted by the sponsor and are presented in Table 9. For Study B201, females in the 3.0 RAD group showed a significantly higher incidence of efficacy failure than female patients in the MMF group. No other statistically significant by-treatment group differences were noted in any of the other subgroups for Study B201. However, conclusions regarding black patients in this study were difficult as the number of subjects falling into that subset was small. For Study B251, no statistically significant by-treatment group differences were observed in any of the age, gender, or racial subsets.

Table 9: Efficacy Failure Subgroup Analyses (ITT Group)						
	Study B201			Study B251		
Subgroup	1.5 RAD (N=194)	3.0 RAD (N=198)	MMF (N=196)	1.5 RAD (N=193)	3.0 RAD (N=194)	MMF (N=196)
Males, n/N (%) RAD vs. MMF*	28/114 (24.6%) p=0.134	29/127 (22.8%) p=0.059	46/139 (33.1%) NA	23/110 (20.9%) p=0.153	34/123 (27.6%) p=0.831	38/132 (28.8%) NA
Females, n/N (%) RAD vs. MMF*	30/80 (37.5%) p=0.159	31/71 (43.7%) p=0.036	15/57 (26.3%) NA	25/83 (30.1%) p=0.490	17/71 (23.9%) p=0.882	16/64 (25.0%) NA
<50 years of age, n/N (%) RAD vs. MMF*	29/115 (25.2%) p=0.307	36/121 (29.8%) p=0.804	35/112 (31.3%) NA	29/129 (22.5%) p=0.208	34/124 (27.4%) p=0.726	37/126 (29.4%) NA
≥50 years of age, n/N (%) RAD vs. MMF*	29/79 (36.7%) p=0.442	24/77 (31.2%) p=0.978	26/84 (31.0%) NA	19/64 (30.1%) p=0.482	17/70 (24.3%) p=1.000	17/70 (24.3%) NA
Black, n/N (%) RAD vs. MMF*	2/4 (50.0%) p=0.517^	3/9 (33.3%) p=0.617^	2/11 (18.2%) NA	11/29 (37.9%) p=0.903	11/36 (30.6%) p=0.610	12/33 (36.4%) NA
Non-Black, n/N (%) RAD vs. MMF*	56/190 (29.5%) p=0.614	57/189 (30.2%) p=0.722	59/185 (31.9%) NA	37/164 (22.6%) p=0.499	40/158 (25.3%) p=0.918	42/163 (25.8%) NA

*Z-test
^Fisher's Exact Test

4.2 Other Special/Subgroup Populations

Subgroup analyses of efficacy failure (i.e., biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) by diabetic status, delayed graft function, and "high risk" status were conducted by the sponsor and are presented in Table 10. No Statistically significant by-treatment group differences were identified for any subgroup in either study. However, conclusions regarding diabetic subjects in Study B201 and subjects with delayed graft function in Study B251 were difficult as the number of subjects falling into those subsets were small.

Table 10: Efficacy Failure Subgroup Analyses (ITT Group)						
	Study B201			Study B251		
Subgroup	1.5 RAD (N=194)	3.0 RAD (N=198)	MMF (N=196)	1.5 RAD (N=193)	3.0 RAD (N=194)	MMF (N=196)
Diabetic, n/N (%) RAD vs. MMF*	4/8 (50.0%) p=0.161^	4/17 (23.5%) p=1.000^	2/12 (16.7%) NA	11/36 (30.6%) p=0.431	12/40 (30.0%) p=0.452	11/48 (22.9%) NA
Non-Diabetic, n/N (%) RAD vs. MMF*	54/186 (29.0%) p=0.517	56/181 (30.9%) p=0.805	59/184 (32.1%) NA	37/157 (23.6%) p=0.275	39/154 (25.3%) p=0.458	43/148 (29.1%) NA
Delayed Graft Func, n/N (%) RAD vs. MMF*	23/45 (51.1%) p=0.805	21/42 (50.0%) p=0.732	21/39 (53.8%) NA	4/15 (26.7%) p=0.109^	7/19 (36.8%) P=0.257^	7/11 (63.64%) NA
No Delayed Graft Func, n/N (%) RAD vs. MMF*	35/149 (23.5%) p=0.684	39/156 (25.0%) p=0.919	40/157 (25.5%) NA	44/178 (24.7%) p=0.878	44/175 (25.1%) p=0.948	47/185 (25.4%) NA
"High Risk", n/N (%) RAD vs. MMF*	41/130 (31.5%) p=0.591	43/135 (31.9%) p=0.637	47/136 (34.6%) NA	24/83 (28.9%) p=0.989	22/80 (27.5%) p=0.858	21/73 (28.8%) NA
"Non-High Risk", n/N (%) RAD vs. MMF*	17/64 (26.6%) p=0.671	17/63 (27.0%) p=0.636	14/60 (23.3%) NA	24/110 (21.8%) p=0.373	29/114 (25.4%) p=0.806	33/123 (26.8%) NA

*Z-test
^Fisher's Exact Test

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The following statistical issues are described within the context of the review. Please see the specified references for details.

- Disproportionate premature treatment discontinuation (Ref: Tables 1 and 4, Figures 2 and 4, Sections 3.1.2 and 3.1.3)
- Partial unblinding at six months (Ref: Reviewer's Comment in Section 3.1.3)
- Unblinding and adjustments in study medication at 12 months (Ref: Section 3.1.4)
- Presentation and critique of primary efficacy analyses (Ref: Tables 3 and 4, Figures 3 and 4, Section 3.1.3)
- Presentation of analyses demonstrating a disproportionate deterioration in renal function in the RAD groups (Ref: Table 6 and 7, Figures 5 and 6, Section 3.2)

5.2 Conclusions and Recommendations

Formulation of overall conclusions regarding the efficacy of 1.5 RAD and 3.0 RAD in comparison to MMF is difficult since many by-treatment group comparisons of both the primary and secondary efficacy endpoints did not afford the luxury of consistent results. In addition sensitivity analyses conducted by this reviewer indicated that many of the primary efficacy results were not robust against the disproportionate premature treatment discontinuation rate.

The protocol prescribed co-primary efficacy analysis (of biopsy proven acute rejection, graft loss, death, or lost-to-follow up within six months) did indicate that each of the RAD groups were noninferior to MMF for both studies. However, not all of these results were robust against the effect of the disproportionate premature treatment discontinuation. When considering premature treatment discontinuation as an efficacy failure (i.e., the modified composite endpoint was biopsy proven acute rejection, graft loss, death, lost-to-follow-up, or premature treatment discontinuation) only the 1.5 RAD dose maintained non-inferiority to MMF for both of the studies. The 3.0 RAD group was not non-inferior to MMF for this endpoint in either study.

The second protocol prescribed co-primary efficacy analysis (of graft loss, death, or lost-to-follow up within 12 months) did indicate that in Study B201 the 1.5 RAD group was non-inferior to the MMF group and in Study B251 the 3.0 RAD group was non-inferior to the MMF group. However, none of these results were robust against the effect of the disproportionate premature treatment discontinuation. When considering premature treatment discontinuation as a failure (i.e., the modified composite endpoint was graft loss, death, lost-to-follow-up, or premature treatment discontinuation) neither the 1.5 RAD or 3.0 RAD groups were non-inferior to MMF in either study and in fact the 3.0 RAD group was statistically significantly worse than the MMF group in both studies.

Efficacy results for the secondary analyses were not consistent across studies. Generally, the results of Study B201 seemed to indicate that the 1.5 RAD dose had more acceptable

efficacy than the 3.0 RAD dose. However, Study B251 seemed to suggest the opposite, that the 3.0 RAD dose had more acceptable efficacy. A summary of the secondary efficacy results previously discussed in this review follows.

- The 1.5 RAD group was non-inferior to the MMF group for both co-primary efficacy endpoints at all time points in Study B201.
- The 1.5 RAD group was *not* non-inferior to the MMF group for the co-primary efficacy endpoint (biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) at 36 months in Study B251. Non-inferiority for this endpoint, dose, and study was achieved for the six and 12 months time points. But results for the graft loss, death, or lost-to-follow-up co-primary composite indicated that the 1.5 RAD group was *not* non-inferior to the MMF group at any of the time points. Examination of each of the components of the composite revealed that the event of graft loss was the primary reason for these results.
- The 3.0 RAD group was non-inferior to the MMF group for the co-primary efficacy endpoint (of biopsy proven acute rejection, graft loss, death, or loss-to-follow-up) at all time points except 36 months for Study B201. But results for the graft loss, death, or lost-to-follow-up co-primary composite indicated that the 3.0 RAD group was *not* non-inferior to the MMF group at 12 months and in fact was statistically significantly *worse* than the MMF group at 36 months. Examination of each of the components of the composite revealed that the event of graft loss was the primary reason for these results.
- The 3.0 RAD group was non-inferior to the MMF group for both the co-primary efficacy endpoints at all time points except 36 months for Study B251.

Statistically significant differences among treatment groups were observed in creatinine and creatinine clearance at numerous time points throughout both studies. In both studies, the 3.0 RAD group had statistically significantly worse median creatinine and creatinine clearance than those of the MMF group beginning at three months and continuing consistently throughout the remainder of the study at 36 months. Except for creatinine in Study B201 at two time points (30 and 36 months), the 1.5 RAD group had statistically significantly worse median creatinine and creatinine clearance when compared to the MMF group beginning at six months and continuing through 36 months.

In light of the significant concerns regarding renal toxicities, the inconsistency of the efficacy results (by dose and study), and the possibility that the incidence of graft loss may increase with the use of RAD, it has been suggested that therapeutic drug monitoring may be a more appropriate method for administering RAD. Exploratory analyses of Studies B201 and B251 considering associations between certain achieved drug concentrations and various efficacy and safety parameters could be conducted. However, it is the assessment of this reviewer that such analyses are not sufficient to make confirmatory conclusions as the subjects have not been randomly assigned to their achieved drug concentrations. Therefore, in the assessment of this reviewer, these studies (or re-analyses of them) cannot be used to justify a safe and effective RAD dose or regimen for the prophylaxis of organ rejection in allogeneic kidney transplant patients.

6. APPENDICES

This section is intentionally left blank.

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

/s/

Ruth Davi
10/16/03 02:04:08 PM
BIOMETRICS

Mohammad Huque
10/16/03 03:14:40 PM
BIOMETRICS

Karen Higgins
10/16/03 03:33:47 PM
BIOMETRICS