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

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA /Serial Number:	21-491
Drug Name:	10 mg Atrasentan
Applicant:	Abbott Laboratories
Indication(s):	Metastatic, Hormore-Refractory Prostate Cancer
Date(s):	Submission Date: December 14, 2004
	Review Completion Date for ODAC: August 15, 2005
Review Priority:	Standard
Biometrics Division:	Division of Biometrics I (HFD-710)
Statistical Reviewer:	Shenghui Tang, Ph.D.
Concurring Reviewer:	Rajeshwari Sridhara, Ph.D., Team Leader
	Kooros Mahjoob, Ph.D., Acting Director
Medical Division:	Oncology Drug Products (HFD-150)
Clinical Team:	Amna Ibrahim, M.D. & John Johnson, M.D.
Project Manager:	Ms. Ann M. Staten
Keywords:	Active control/superiority, log-rank test, Cox regression

Table of Contents

1 Executive Summary

This is a review of NDA21-491/N-000 for the use of the drug Xinlay (trade name Atrasentan, 10mg, orally administered) in the treatment of male subjects diagnosed with metastatic, hormone-rcfractory prostate cancer (HRPC).

The sponsor has submitted results of the final analyses from the M00-211 and M96-594 studies designed to evaluate the efficacy of 10 mg Atrasentan for the treatment of metastatic HRPC. M00-211 was a phase 3, randomized, double-blind, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. Study M96-594 was a Phase 2, randomized, double-blind, placebo-controlled, parallel-group, multicenter study of Atrasentan.

Because of difference in definitions of time-to-disease-progression (TDP) between the two studies, this review will focus on the pivotal phase III study M00-211 with a brief review of the efficacy results from Study M96-594.

This application will be discussed at the Oncology Advisory Committee meeting on September 13, 2005

1.1 Conclusions and Recommendations

In this reviewer's opinion the results from both studies failed to demonstrate efficacy with respect to a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group.

After the studies failed to demonstrate the efficacy, and after the submission of NDA to the Agency, the applicant is seeking approval based on a post-hoc, subgroup, exploratory analysis in the subgroup of patients with bone metastases at baseline in the study M00-211 and claims a favorable effect in this subpopulation (per the applicant, HR=0.813, CI= 0.685-0.965; p-value= 0.016 (unadjusted)).

The study M00-211 was designed to answer a question about the overall Atrasentan effect in the entire population, not to answer questions about the subgroups. The statistical plan was never amended to include baseline bone metastasis sub-group analysis as primary efficacy analyses. The analysis for patients with bone metastases at baseline was not pre-specified. Hence, the observed TDP difference in this subgroup is considered as an exploratory and hypothesis generating analysis. The p-value from this analysis is not interpretable since all the type I error rate has been spent in the failed primary protocol pre-specified analysis. Furthermore, this data has been analyzed multiple times and within multiple subgroups with no type I error adjustment. The findings from the subgroup analyses should be confirmed through other studies.

1.2 Brief Overview of Clinical Studies

By submitting this NDA application, the sponsor is seeking approval of using Xinlay (trade name Atrasentan, 10mg, orally administered) in the treatment of male subjects diagnosed with metastatic, hormone-rcfractory prostate cancer.

M00-211 was a phase 3, randomized, double-blinded, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. The men participating in this study were diagnosed with hormone-refractory prostate cancer that had been treated with surgical and/or chemical castration and were progressing despite androgen suppression at the time of study entry as demonstrated by a rising PSA. These men had evidence of distant metastases. A total of 809 male subjects were enrolled at 179 investigative sites and were randomly assigned to receive either Atrasentan or placebo.

Study M96-594 was a Phase 2, randomized, double-blind, placebo-controlled, parallel-group, multicenter study of Atrasentan. This study consisted of a 14-day screening period and a double-blind treatment period that continued unti1 the subject experienced clinica1 disease progression or otherwise discontinued from the study. At least 204 subjects were to be enrolled to ensure 68 subjects per treatment arm. Subjects were randomized to receive: 10 mg Atrasentan, 2.5 mg Atrasentan, or placebo daily in a 1:1:1 ratio, in addition to their standard care. A subject was considered to have completed the study after presenting evidence of disease progression (event).

1.3 Statistical Issues and Findings

For study M00-211, there were 311 events (77.6%) for disease progression in the placebo arm and 299 events (73.3%) in the 10 mg Atrasentan arm. Among the 311 events in the placebo arm, there were 98 events from USA. In the treatment arm, 100 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the distribution of the time-to-disease progression for the placebo was not significantly different from the distribution for the 10 mg Atrasentan treatment group (p=0.143). An unstratified log-rank test showed a similar result. The hazard ratio for the time-to-disease progression (TDP) in the 10 mg Atrasentan arm, as compared with the placebo arm, was 0.881.

For study M96-594, an unstratified log-rank test also showed that the distribution of the time-to-disease progression for the placebo was not significantly different from the distribution for the 10 mg Atrasentan treatment group (p=0.1323 without adjustment for multiple comparisons). The hazard ratio for the time-to-disease progression (TDP) in the 10 mg Atrasentan arm, as compared with the Placebo arm, was 0.769.

The primary TDP analyses in both studies failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group. All the type I error rate has been spent in the failed primary protocol pre-specified analyses.

Statistical Issues:

1. The independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint using the $G^{1,1}$ analysis on the intent-to-treat population. The decision was based on 809 subjects, 343 of which experienced disease progression. The last subject's last dose of administration of the study drug was on 19 March 2003.

2. The primary TDP analysis in M00-211 failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group ($p = 0.143$). This analysis used all of the two-sided alpha of 0.05.

   Per the sponsor specified protocol:
   "If the primary efficacy analysis is statistically significant at the $\alpha=0.05$ level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule ……. If the primary efficacy analysis is not statistically significant at the $\alpha=0.05$ level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values."

   With the failed primary analysis, all pre-specified secondary and tertiary analyses were considered as exploratory / hypothesis generating.

3. Study M00-211 was designed to answer a question about the overall Atrasentan effect in the entire population, not to answer questions about the subgroups. The protocol or the statistical plan was never amended to include per-protocol analysis and baseline bone metastasis sub-group analysis as primary efficacy analyses. Furthermore, the analysis for patients with bone metastases at baseline was not pre-specified. Although the per-protocol analysis was outlined in the protocol, it was considered as tertiary. The protocol further indicated that significance for the subgroup analyses will not be declared, regardless of the observed p-values.

4. No statistical adjustment was made for the multiple analyses (subgroup analyses) and multiple hypotheses . Without any type I error adjustment for the

multiple comparisons, these analyses are not interpretable. The findings from the analyses should be confirmed through other studies.

5. According to the ICH E9 guidelines (Statistical Principles for Clinical Trials), any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted. The ICH E9 also states that "…., an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol".

   Subgroup analyses usually have high false positive or false negative rate. Statistical results should be interpreted with extreme caution because false positive findings may increase as the number of significance tests performed increases.

6. It is impossible to correctly adjust the nominal p-value for multiple comparisons post hoc. Post hoc analyses are considered as hypothesis generating. Subgroup analyses suggest hypotheses worth examining in other studies. The strength of evidence for efficacy is discredited with multiple subgroup analyses where one could have many chances to find a difference between two arms. With no pre-specified analyses, the Type I error will also be inflated.

7. The primary ITT analysis of time-to-disease progression was pre-specified in the M002-11. However, the M00211 did not pre-specify any adjustment procedure for the subgroup analyses using the per-protocol population or baseline bone metastatic patient population. According to the ICH E9, adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan. Without any allocation of type I error for the analyses using the per-protocol population or baseline bone metastatic patient populations, the results from the subgroup analyses will be considered as supportive to the ITT analysis and may not be claimed in the label.

8. M96-594 is a smaller dose ranging study, which included three study arms: placebo, 2.5 mg Atrasentan, and 10 mg Atrasentan. The design is different from the design used in the phase 3 study M00-211.

9. M96-594 used a different time-to-disease-progression (TDP) definition from the study M00-211 (Table 8). No independent review of progression evaluation was conducted in study M96-594. The primary TDP analysis in this study also failed to demonstrate a delay in disease progression. Therefore, its TDP results cannot be used as supportive evidence to the study M00-211.

10. M96-594 was a phase 2 study which randomized 3 arms (placebo, 2.5mg and 10mg of Astrasentan); M000-211 was a phase 3 study which had two arms (10mg Astrasentan and placebo). This reviewer does not believe in the demonstration of efficacy based on results from pooling trials together, especially when (a) neither of the trials individually showed a statistically significant difference; (b) both studies had different definitions of TDP and no independent review of progression evaluation was conducted in study M96-594; (c) the proposed analysis for pooling trials together is a post-hoc analysis; (d) for the pooled analysis, it is not clear how type I error is controlled.

11. Efficacy demonstration should be solely based on results of the primary analysis from individual trials, where the primary analysis is pre-specified and agreed upon by the Agency. Efficacy claims based on results from any other analyses (such as pre-specified exploratory or post-hoc) can only inflate the false positive error rate and may not be considered for regulatory approval.

12. Quality of Life (QoL) was defined as a tertiary analysis in M00-211. QoL was assessed using two scales: the Functional Assessment of Cancer Therapy – Prostate (FACT-P) and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). The statistical plan was never amended to include QOL as primary efficacy analysis. No statistical adjustment was made for the multiple QoL analyses.

    The sponsor's study report stated that *in the ITT analysis, Atrasentan treatment resulted in a difference in QoL in favor of Atrasentan as measured by the disease-specific PCS score (P = 0.032).* This reviewer questions this statement because of the following reasons:

    1) The protocol did not pre-specify the statistical hypothesis, particularly the alternative hypothesis, which was used in the testing procedure. Hence, we do not know whether the PCS mean change is meaningful. Also with the PCS scores ranging from a possible 0 to 48, it is difficult to interpret the observed PCS mean change of 1.02.
    2) Due to missing values, the PCS analysis did not include all patients. There were 32 patients in placebo arm who had missing values; 43 patients in the Atrasentan arm. Therefore, the analysis was not based on the ITT population.
    3) Several statistical tests were performed without any statistical adjustment for the multiple analyses. The p-value of 0.032 is not interpretable.
    4) Both QoL instruments ask the patients to rate the symptom based on their experience over the past week.

Findings:

Both studies M00-211 and M96-594 failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group (Table 1 and Table 2).

Table 1. Primary Efficacy TDP Analysis in ITT Population (M00-211)

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	408	401
Number of events (%)	299 (73.3%)	311 (77.6%)
Median (days), 95%CI	91 (86, 97)	86 (85, 88)
Stratified $G^{1,1}$ test	P=0.143	
Unstratified Logrank test	P=0.123	
Hazard ratio (95% CI)[1]	0.881 (0.751, 1.033)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the Placebo arm.

Table 2. Primary Efficacy TDP Analysis in ITT Population (M96-594)

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	89	104
Number of events (%)	58 (65.2%)	77 (74.0%)
Median (days), 95% CI	183 (132, 225)	137 (116, 167)
Unstratified Logrank test	p=0.1323*	
Hazard ratio (95% CI)[1]	0.769 (0.545, 1.085)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the Placebo arm.
*Without adjustment for multiple comparisons

2 Introduction

2.1 Overview

The sponsor is seeking approval of using Xinlay (trade name Atrasentan, 10mg, orally administered) in the treatment of male subjects diagnosed with metastatic, hormone-refractory prostate cancer.

Adenocarcinoma of the prostate is the most commonly diagnosed non-cutaneous cancer and is the second most common cause of cancer-related death in men, second only to lung cancer. Globally, over 200,000 patients die from metastatic prostate cancer every year, with over 35,000 in the United States and 80,000 in Europe. If completely confined to the prostate gland, prostate cancer can be cured by definitive local therapy with radical prostatectomy or radiotherapy. Many patients, however, will relapse following primary therapy, and will be treated with androgen-deprivation therapy (ADT), a concept established in 1941 by Huggins. ADT is effective because nearly all prostate cancers are highly dependent upon androgens for growth, and interruption of this stimulus causes prolonged growth arrest. Unfortunately, in most men, after several years of ADT, the prostate cancer becomes androgen-refractory, progressing despite castrate levels of serum androgens. This form of the disease is termed hormone-refractory prostate cancer (HRPCa) and is a fatal condition with no approved therapeutic option apart from palliation. It is associated with the development of painful bony metastases preceding death, which occurs 12 to 24 months after HRPCa onset.

Recent studies have presented consistent evidence that a selective ET_AR antagonist would provide therapeutic benefit for HRPCa patients, particularly at the bone/tumor interface. Atrasentan (A-147627; ABT-627) is a selective, potent, orally active ET_AR antagonist as a therapy for PCa. The combination of Taxotere and Prednisone was recently approved based on improved overall survival in this setting. Mitoxantrone plus Prednisone was approved based on improvement in pain symptom.

The efficacy of 10 mg Atrasentan for the treatment of metastatic HRPC was evaluated in the following studies: M00-211, M96-594 and M96-500.

M00-211 was a phase 3, randomized, double-blind, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. Study M96-594 was a phase 2, multicenter, multinational, randomized, double-blind, placebo-controlled, parallel-group study of orally administered 2.5 mg or 10 mg Atrasentan once daily (QD) versus placebo QD. Both studies evaluated time to disease progression as a primary endpoint. However, the studies had different definitions of TDP. This review will focus on the pivotal phase 3 study M00-211 with a brief review of the efficacy results from Study M96-594.

Study M96-500, which was designed to treat subjects for 84 days, included a substantially more advanced symptomatic metastatic HRPC patient population with prostate cancer–related pain requiring opiate analgesia at baseline. This study did not evaluate time to disease progression; therefore, it is not included in this review.

2.1.1 Background

Study M00-211

M00-211 was a phase 3, randomized, double-blind, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. The men participating in this study were diagnosed with hormone-refractory prostate cancer that had been treated with surgical and/or chemical castration and was progressing despite androgen suppression at the time of study entry as demonstrated by a rising PSA. These men had evidence of distant metastases.

A total of 809 male subjects were enrolled at 179 investigative sites and were randomly assigned to receive either Atrasentan or placebo.

Simulations performed to determine the power for the primary analysis at the two-sided 0.05 significance level indicated that 650 events of disease progression yield 90% power. To deliver 650 events of disease progression within an acceptable timeframe, between 900 and 1000 subjects needed to be enrolled. **The independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized.** Of these, all 809 were included in the efficacy analyses and 801 were included in the safety analyses (eight subjects did not receive study drug and are not included in the safety analyses). The last subject's last dose of administration of the study drug was on 19 March 2003. **The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint using the G1,1 analysis on the intent-to-treat population.** The decision was based on 809 subjects, 343 of which experienced disease progression. The study blind for M00-211 was broken on 16 May 2003. There were, therefore, 610 events of disease progression instead of the 650 anticipated events.

The first subject's first dose of study drug was administrated on 25 June 2001 and the last subject's last dose of study drug was administrated on 19 March 2003.

Duration of Treatment: Subjects were considered to have completed the study once they experienced an event of disease progression, at which time they were eligible to enroll in the open-label extension study, M00-258. Subjects who

remained active but did not experience disease progression at the time the blind was broken were also eligible to enter study M00-258. The mean number of days of exposure to Atrasentan was 144 (standard deviation: 91, median: 119 days, range: 2 to 532 days). A total of 354 subjects (87.6% of all Atrasentan-treated subjects) were exposed to Atrasentan for at least 3 months, 161 (39.9%) for at least 6 months, 63 (15.6%) for at least 9 months, and 21 (5.2%) for at least one year.

Study M96-594

Study M96-594 was a phase 2, multicenter, multinational, randomized, double-blind, placebo-controlled, parallel-group study of orally administered 2.5 mg or 10 mg Atrasentan once daily (QD) versus placebo QD (See Section 3.1.2).

2.1.2 Statistical Issues

1. The independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint using the $G^{1,1}$ analysis on the intent-to-treat population. The decision was based on 809 subjects, 343 of which experienced disease progression. The last subject's last dose of administration of the study drug was on 19 March 2003.

2. The primary TDP analysis in M00-211 failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group ($p = 0.143$). This analysis used all of the two-sided alpha of 0.05.

   Per the sponsor specified protocol:
   "If the primary efficacy analysis is statistically significant at the α=0.05 level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule ……. If the primary efficacy analysis is not statistically significant at the α=0.05 level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values."

   With the failed primary analysis, all pre-specified secondary and tertiary analyses were considered as exploratory / hypothesis generating.

3. Study M00-211 was designed to answer a question about the overall Atrasentan effect in the entire population, not to answer questions about the subgroups. The protocol or the statistical plan was never amended to include per-protocol analysis and baseline bone metastasis sub-group analysis as primary efficacy

analyses. Furthermore, the analysis for patients with bone metastases at baseline was not pre-specified. Although the per-protocol analysis was outlined in the protocol, it was considered as tertiary. The protocol further indicated that significance for the subgroup analyses will not be declared, regardless of the observed p-values.

4. No statistical adjustment was made for the multiple analyses (subgroup analyses) and multiple hypotheses . Without any type I error adjustment for the multiple comparisons, these analyses are not interpretable. The findings from the analyses should be confirmed through other studies.

5. According to the ICH E9 guidelines (Statistical Principles for Clinical Trials), any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted. The ICH E9 also states that "…., an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol".

   Subgroup analyses usually have high false positive or false negative rate. Statistical results should be interpreted with extreme caution because false positive findings may increase as the number of significance tests performed increases.

6. It is impossible to correctly adjust the nominal p-value for multiple comparisons post hoc. Post hoc analyses are considered as hypothesis generating. Subgroup analyses suggest hypotheses worth examining in other studies. The strength of evidence for efficacy is discredited with multiple subgroup analyses where one could have many chances to find a difference between two arms. With no pre-specified analyses, the Type I error will also be inflated.

7. The primary ITT analysis of time-to-disease progression was pre-specified in the M002-11. However, the M00211 did not pre-specify any adjustment procedure for the subgroup analyses using the per-protocol population or baseline bone metastatic patient population. According to the ICH E9, adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan. Without any allocation of type I error for the analyses using the per-protocol population or baseline bone metastatic patient populations, the results from the subgroup analyses will be considered as supportive to the ITT analysis and may not be claimed in the label.

8. M96-594 is a smaller dose ranging study, which included three study arms: placebo, 2.5 mg Atrasentan, and 10 mg Atrasentan. The design is different from

the design used in the phase 3 study M00-211.

9. M96-594 used a different time-to-disease-progression (TDP) definition from the study M00-211 (Table 8). No independent review of progression evaluation was conducted in study M96-594. The primary TDP analysis in this study also failed to demonstrate a delay in disease progression. Therefore, its TDP results cannot be used as supportive evidence to the study M00-211.

10. M96-594 was a phase 2 study which randomized 3 arms (placebo, 2.5mg and 10mg of Astrasentan); M000-211 was a phase 3 study which had two arms (10mg Astrasentan and placebo). This reviewer does not believe in the demonstration of efficacy based on results from pooling trials together, especially when (a) neither of the trials individually showed a statistically significant difference; (b) both studies had different definitions of TDP and no independent review of progression evaluation was conducted in study M96-594; (c) the proposed analysis for pooling trials together is a post-hoc analysis; (d) for the pooled analysis, it is not clear how type I error is controlled.

11. Efficacy demonstration should be solely based on results of the primary analysis from individual trials, where the primary analysis is pre-specified and agreed upon by the Agency. Efficacy claims based on results from any other analyses (such as pre-specified exploratory or post-hoc) can only inflate the false positive error rate and may not be considered for regulatory approval.

12. Quality of Life (QoL) was defined as a tertiary analysis in M00-211. QoL was assessed using two scales: the Functional Assessment of Cancer Therapy – Prostate (FACT-P) and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). The statistical plan was never amended to include QOL as primary efficacy analysis. No statistical adjustment was made for the multiple QoL analyses.

    The sponsor's study report stated that *in the ITT analysis, Atrasentan treatment resulted in a difference in QoL in favor of Atrasentan as measured by the disease-specific PCS score (P = 0.032).* This reviewer questions this statement because of the following reasons:

    1) The protocol did not pre-specify the statistical hypothesis, particularly the alternative hypothesis, which was used in the testing procedure. Hence, we do not know whether the PCS mean change is meaningful. Also with the PCS scores ranging from a possible 0 to 48, it is difficult to interpret the observed PCS mean change of 1.02.
    2) Due to missing values, the PCS analysis did not include all patients. There were 32 patients in placebo arm who had missing values; 43

patients in the Atrasentan arm. Therefore, the analysis was not based on the ITT population.

3) Several statistical tests were performed without any statistical adjustment for the multiple analyses. The p-value of 0.032 is not interpretable.
4) Both QoL instruments ask the patients to rate the symptom based on their experience over the past week.

2.2 Data Sources

Data and electronic documents used for this review are located on the network with path "\\Cdsesub1\n21491\N_000\2005-02-24" and " \\Cdsesub1\n21491\N_000\2004-12-13" in the EDR.

3 Statistical Evaluation

3.1 Evaluation of Efficacy

The sponsor has submitted results of the final analysis from the M00-211 and M96-594 studies designed to evaluate the efficacy of 10 mg Atrasentan for the treatment of metastatic HRPC.

This review will focus on the pivotal phase 3 study M00-211 with a brief review of the efficacy results from Study M96-594.

3.1.1 Study M00-211

M00-211 was a phase 3, randomized, double-blind, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. The first subject's first dose of study drug was on 25 June 2001 and the last subject's last dose of study drug was on 19 March 2003.

3.1.1.1 Study Design

A total of 809 male subjects were enrolled at 179 investigative sites and were randomly assigned to receive either Atrasentan or placebo.

According the protocol, all screening procedures were to be performed after written informed consent was obtained and within 35 days prior to randomization.

On Day 1, subjects who met the enrollment criteria were randomly assigned in a 1:1 ratio to receive either 10 mg Atrasentan or matching placebo. Subjects were assigned a 4-digit subject number and were given study drug prior to leaving the clinic. During the course of treatment, the subjects visited the study site on Day 14, on Weeks 4, 8, and 12, and every 6 weeks thereafter. At each visit, subjects

were assessed for safety and clinical evidence of disease progression. Study drug was dispensed at Day 1, at Weeks 4, 8, and 12, and every 6 weeks thereafter. Every 12 weeks, subjects were evaluated for disease progression by radiographic imaging. If a subject experienced symptoms suspected to be related to disease progression, an appropriate radiographic scan may have been performed prior to the scheduled 12-week radiographic scan. Only the same type scan as the baseline scan could be used for purposes of documenting disease progression.

A subject was considered to have completed the double-blind treatment period once the principal investigator received notification from Abbott Laboratories that the event of disease progression had been confirmed by the independent reviewer or if the subject was active in the trial when the double-blind treatment period ended. Subjects who did not complete the study were classified as having discontinued from the study prematurely. Subjects were to have a final assessment (final visit) upon study completion or premature discontinuation from the study. Subjects who completed the study were eligible to participate in the open-label extension study.

The double-blind treatment period for this study was to end once 650 subjects had experienced disease progression. Subjects who elected to discontinue study drug prior to experiencing an event of disease progression could remain in the study and follow a similar schedule of assessments, regardless of subsequent therapies, in order to be assessed for disease progression. Subjects who did not enter the extension study returned for a safety evaluation 30 days after their final visit. Subjects were assessed for post-treatment survival at 3-month intervals after the last study visit.

A schematic of the study design is shown in Figure 1.

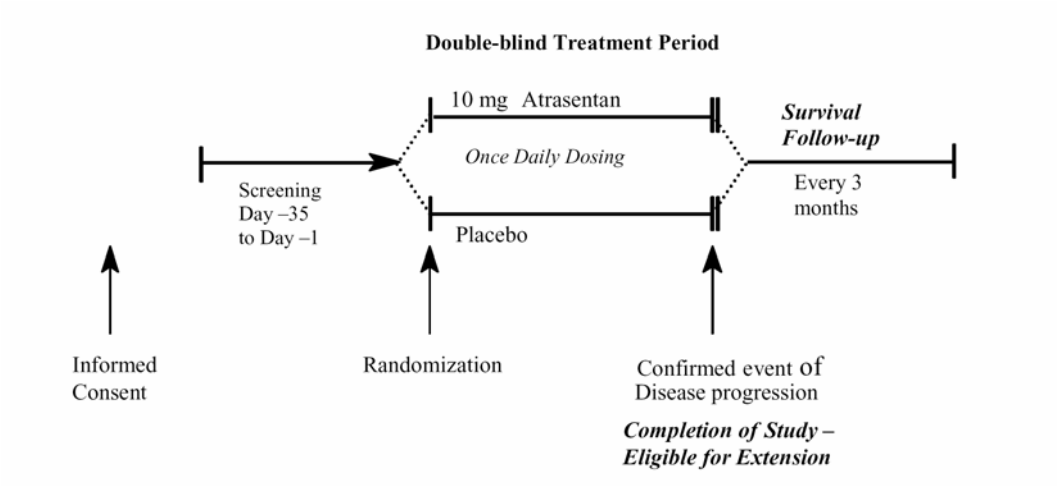

Figure 1. Study Design Schematic (M00-211)

Reviewer's Comments:

The independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. Of these, all 809 were included in the efficacy analyses and 801 were included in the safety analyses (eight subjects did not receive study drug and are not included in the safety analyses). Upon the recommendation of the IDMC, the study was stopped on 10 February 2003. **The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint** using the $G^{1,1}$ analysis on the intent-to-treat population. The decision was based on 809 subjects 343 of which experienced disease progression. Investigators were notified, and all active subjects were to discontinue the study within 4 weeks.

3.1.1.2 Study Objectives

The objectives of Study M00-211 were to evaluate the safety and efficacy of orally administered 10 mg Atrasentan compared with placebo in the treatment of male subjects diagnosed with metastatic, hormone-refractory prostate cancer. Efficacy was measured by time to disease progression. The secondary objective of this study was to evaluate the effect of 10 mg Atrasentan on the following endpoints: biochemical bone markers (specifically bone alkaline phosphatase), time to PSA progression, bone scan index, and survival.

3.1.1.3 Efficacy Endpoints

According to the protocol, the primary efficacy assessment of time to disease progression was determined by the time from randomization to the onset of the earliest of the following events:

1) Pain due to prostate cancer requiring one or more of the following palliative interventions, defined as:
   opioid therapy: a) intravenous, intramuscular, or subcutaneous opioid therapy administered as a single dose; b) oral or transdermal opioid analgesic use administered for 10 out of 14 consecutive days
   glucocorticoid therapy: a) initiation of >=5 mg oral prednisone (or equivalent) for 10 out of 14 consecutive days for subjects not currently on oral steroids b) doubling of the subject's current chronic steroid therapy for 10 out of 14 consecutive days for subjects on a stable dose of oral steroids
   radionuclide therapy
   radiation therapy
   chemotherapy

Evidence of disease at the site of pain was required. Pain requiring only non-opioid analgesics was not considered disease progression.

2) A skeletal-related event — a pathologic or vertebral compression fracture not related to trauma, prophylactic radiation, or surgery for an impending fracture, or spinal cord compression. Evidence of disease at the site is required.
3) An event due to metastatic prostate cancer requiring intervention, e.g., urinary tract obstruction, malignant pleural effusion, brain metastases, or other similar events. Evidence of disease at the site was required. An increase in PSA was not considered an event of disease progression.
4) One bone scan subsequent to baseline demonstrating two or more new skeletal lesions (this specific criterion was developed after consultation with the FDA). An increase in size or intensity of known skeletal lesions was not considered disease progression.
5) One CT or MRI scan subsequent to baseline demonstrating evidence of extra-skeletal disease progression according to a modified Response Evaluation Criteria in Solid Tumors (RECIST) criteria.

   a) an increase in the sum of the longest diameters of target lesions (measuring >=2 cm in longest diameter on baseline scan) by >=20% when compared with the smallest sum of the longest diameters of these target lesions
   b) an increase in size of a solitary sub-target lesion (measuring ≥1.5 cm but <2.0 cm in longest diameter on the baseline scan) to ≥2.4 cm in longest diameter
   c) unequivocal progression of existing lesions not identified as target lesions as determined by an independent reviewer
   d) the appearance of one or more new extra-skeletal lesions (≥1.5 cm, unidimensional) consistent with prostate cancer.

Disease progression was determined only by comparing images generated using the same technique, i.e., bone scan to CT scan changes were not acceptable. All events and dates of progression were reviewed and confirmed by an independent reviewer. The confirmed event with the corresponding earliest date was used as the primary endpoint.

Secondary Efficacy Endpoints included:

1) Biochemical bone markers, specifically bone alkaline phosphatase
2) Time to PSA progression
3) Bone Scan Index (BSI)
4) Survival

Other objectives included evaluating the effect of 10 mg Atrasentan on quality of life (QoL) and performance status. Quality of life was assessed during this study

using two validated scales: the Functional Assessment of Cancer Therapy – Prostate (FACT-P) and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). In addition, population pharmacokinetic analyses were performed.

Reviewer's Comment:

This review will focus on the primary time to disease progression efficacy analysis and the four secondary analyses listed above.

3.1.1.4 Sample Size Considerations

According to the protocol, simulations performed to determine the power for the primary analysis at the two-sided 0.05 significance level indicated that 650 events of disease progression would yield 90% power. To deliver 650 events of disease progression within an acceptable timeframe, between 900 and 1000 subjects needed to be enrolled. According the study report, the independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. Of these, all 809 were included in the efficacy analyses and 801 were included in the safety analyses (eight subjects did not receive study drug and were not included in the safety analyses). The last subject's last dose of administration of the study drug was on 19 March 2003. The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint using the G1,1 analysis on the intent-to-treat population. The decision was based on 809 subjects, 343 of which experienced disease progression. The study blind for M00-211 was broken on 16 May 2003. In this submission there were 610 events of disease progression instead of the 650 anticipated events.

The first subject's first dose of administration of the study drug was on 25 June 2001 and the last subject's last dose of administration of the study drug was on 19 March 2003.

3.1.1.5 Efficacy Analysis Methods

Primary Analysis of Efficacy

The protocol stated that the distribution of time-to-disease progression will be estimated for each treatment group using Kaplan-Meier methodology. A weighted log rank statistic stratified by region (US sites vs. non-US sites) will be used to test the null hypothesis that the distribution of the time-to-disease progression for the placebo and the 10 mg Atrasentan treatment groups are the same. The weighted log rank test statistic, $G^{1,1}$, is a member of the class of weighted log rank statistics, $G^{\rho,\gamma}$. The $G^{\rho,\gamma}$ class of statistics also includes the standard log rank

($G^{0,0}$) and Prentice-Wilcoxon ($G^{1,0}$) statistics. The proportional hazards analysis model was also applied to calculate hazard ratios.

The protocol further stated that, if v_j is the $G^{1,1}$ statistic (for comparing two treatment groups) fram stratum j and V_j is the variance of the $G^{1,1}$ statistic from stratum j then the stratified $G^{1,1}$ test statistic for two strata (j=1,2) is $(v_1 + v_2)^2/(V_1 + V_2)$.

According the protocol, for a given subject, time-to-disease progression will be defined as the number of days from the day the subject was randomized to the day the subject experiences a confirmed event of disease progression. All events of disease progression, as confirmed by the Independent Reviewer, will be included, regardless of whether the event occurred while the subject was still taking study drug or had previously discontinued study drug. If the subject does not have a confirmed event of disease progression, the subject's data will be censored at the date of the subject's last available evaluation. This date will be the date of the last available vital sign measurement, performance status assessment, or physical exam. An exception to this rule may occur if the subject has an unconfirmed event of disease progression (Disease Progression Packet submitted but an event of disease progression was not confirmed by Independent Reviewer). In this case, the last available date of evaluation will be the date of disease progression determined by the investigator for the unconfirmed event or the date of the last available vital sign measurement, performance status assessment, or physical exam, whichever is last.

Secondary Analyses of Efficacy

The study protocol pre-specified the following step-down rule to adjust multiple comparisons:

If the primary efficacy analysis is statistically significant at the α=0.05 level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule, with analyses performed in the following order: (1) mean change from baseline to finial value in bone alkaline phosphatase, (2) time to onset of PSA progression, (3) mean rate of change from baseline to final value in total bone scan index, and (4) survival. If any of these secondary analyses does not achieve statistical significance at the α=0.05 level, then statistical significance will not be declared for the subsequent secondary analyses, regardless of the observed p-values. **If the primary efficacy analysis is not statistically significant at the α=0.05 level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values.**

Mean change from baseline to final value in bone alkaline phosphatase will be calculated for each treament group and compared using an analysis of covariance (ANCOVA) with treatment group and baseline bone alka1ine phosphatase value as the factors. Subjects lacking either a baseline or a final value for bone alkaline phosphatase were not included in this analysis.

The distribution of the time to onset of PSA progression was estimated for each treatment group using Kaplan-Meier methodology. The stratified $G^{1,1}$ test was used to compare the time to onset of PSA progression between 10 mg Atrasentan and placebo.

Mean rate of change in total bone scan index was calculated for each treatment group and compared using an ANCOVA analysis, with treatment group and baseline value of total bone scan index as the factors. Subjects lacking either a baseline or a final value for total bone scan index were not included in this analysis.

The distribution of the time to death was estimated for each treatment group using Kaplan-Meier methodology. The stratified $G^{1,1}$ test was used to compare the time to death between 10 mg Atrasentan and placebo.

Reviewer's Comments:

The protocol or statistical plan did not to include per-protocol analysis (as primary efficacy analysis), analyses on patients with bone metastasis sub-group, or QOL. No statistical adjustment was planned for these multiple analyses.

3.1.1.6 Sponsor's Results and Statistical Reviewer's Findings/ Comments

According the study report, the independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. Of these, all 809 were included in the efficacy analyses. The first subject's first dose of study drug was on 25 June 2001 and the last subject's last dose of study drug was on 19 March 2003.

Data were analyzed for an intent-to-treat population of 809 subjects, which included all data up to the date when all subjects had completed or discontinued from the study. This included 8 subjects randomized into the study but never treated. The primary analysis set consists of data collected and adjudicated from study M00-211 before the 16 May 2003 blind-break and survival data were included through a cutoff of 30 April 2004.

3.1.1.6.1 Baseline Characteristics

The baseline Characteristics of the overall population are presented in Table 3.

Table 3. Baseline Characteristics of the Patients (M00-211)

Characteristic	10 mg Atrasentan (N=408)	Placebo (N=401)	ALL (N=809)
Age — yr			
Mean (SE)	72.3 (0.40)	71.3 (0.41)	71.8 (0.29)
Median (Range)	73 (45–93)	72 (45–92)	72 (45-93)
Age grouped — no. (%)			
<65	68 (16.7)	78 (19.5)	146 (18.0)
+65	340 (83.3)	323 (80.5)	663 (82.0)
Race — no. (%)			
Caucasian	384 (94.1)	386 (96.3)	770 (95.2)
Black	18 (4.4)	8 (1.9)	26 (3.2)
Oriental/Asian	4 (1.0)	4 (1.0)	8 (1.0)
Others	2 (0.5)	3 (0.7)	5 (0.6)
Weight — kg			
Mean (SE)	83.9 (0.71)	85.2 (0.78)	84.5 (0.53)
Median (Range)	81.2 (53.5–176.9)	83.0 (47.0–154.2)	82.1 (47.0-176.9)
Hight — cm			
Mean (SE)	174.4 (0.38)	85.2 (0.78)	174.6 (0.27)
Median (Range)	175.0(152.0–198.1)	175.0(152.4–195.0)	175.0(152.0-198.1)
Karnofsky performance-status score — no. (%)			
≤70	10 (2.5)	12 (3.0)	22 (2.7)
80	40 (9.8)	41 (10.2)	81 (10.0)
90	151 (37.0)	125 (31.2)	276(34.1)
100	207 (50.7)	223 (55.6)	430(53.2)
Hemoglobin — g/dL			
Mean (SE)	13.3 (0.06)	13.1 (0.07)	13.2 (0.05)
Median (Range)	13.4 (9.3–17.4)	13.2 (9.1–18.1)	13.3 (9.1-18.1)
LDH — IU/L			
Mean (SE)	200.7 (4.28)	221.9 (8.61)	211.2 (4.79)
Median (Range)	186.0 (97.0–1318.0)	188.0(108.0–2365.0)	186.0 (97.0-2365.0)
Total alkaline phosphatase — IU/L			
Mean (SE)	196.8 (20.68)	208.2 (17.27)	202.4 (13.49)
Median (Range)	110.0 (36.0–5482.0)	112.0 (41.0–3774.0)	110.0 (36.0-5482.0)
Bone alkaline phosphatase — ng/mL			
Mean (SE)	58.6 (7.70)	59.7 (6.27)	59.1 (4.97)
Median (Range)	25.5 (2.0–1903.8)	24.8 (2.0–1599.0)	25.2 (2.0-1903.8)
PSA — ng/mL			
Mean (SE)	212.2 (24.0)	218.1 (21.3)	215.1 (16.1)
Median (Range)	69.8 (1.7–5784.0)	79.6 (2.2–5424.8)	72.9 (1.7-5784.0)
Screening testosterone — ng/dL			
Mean (SE)	12.7 (0.35)	12.7 (0.36)	12.7 (0.25)
Median (Range)	11.5 (2.9–57.7)	11.3 (2.9–46.0)	11.5 (2.9-57.7)
Total Gleason Score			
Mean (SE)	7.2 (0.08)	7.3 (0.09)	7.3 (0.06)
Median (Range)	7.0 (3.0–10.0)	7.0 (2.0–10.0)	7.0 (2.0-10.0)
Bone Scan Index			
Mean (SE)	4.3 (0.37)	5.5 (0.46)	4.9 (0.29)
Median (Range)	1.4 (0.0–47.9)	1.8 (0.0–44.5)	1.5 (0.0-47.9)
Time since diagnosis — yrs			
Mean (SE)	5.8 (0.19)	5.5 (0.18)	5.7 (0.13)
Median (Range)	5.0 (0.3–23.7)	4.8 (0.1–23.2)	4.9 (0.1-23.7)

Reviewer's Comments:

In the overall patient population the baseline characteristics appear to be balanced between the two treatment arms.

3.1.1.6.2 Primary Efficacy Analyses

According to the study protocol, the primary efficacy analysis in this submission is time to disease progression analysis in the Intention-to-Treat population. As specified in the protocol, a weighted log-rank test stratified by region (US sites vs. non-US sites) was performed to compare time to disease progression (TDP) between the 10 mg Atrasentan arm and the placebo arm in the ITT population.

There were 311 events (77.6%) for disease progression in the placebo arm and 299 events (73.3%) in the 10 mg Atrasentan arm. Among the 311 events in the placebo arm, there were 98 events from USA. In the treatment arm, 100 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the distribution of the time-to-disease progression for the placebo was not significantly different from and the distribution for the 10 mg Atrasentan treatment group (p=0.143). The hazard ratio for the time-to-disease progression in the 10 mg Atrasentan arm, as compared with the Placebo arm, was 0.881.

The results from the stratified $G^{1,1}$ test and unstratified log-rank test are presented in the Table 4. The Kaplan-Meier curves for the ITT population are illustrated in Figure 2.

Table 4. Primary Efficacy TDP Analysis in ITT Population

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	408	401
Number of events (%)	299 (73.3%)	311 (77.6%)
Median (days), 95% CI	91 (86, 97)	86 (85, 88)
Stratified $G^{1,1}$ test	p=0.143	
Unstratified Logrank test	p=0.123	
Hazard ratio (95% CI)[1]	0.881 (0.751, 1.033)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the placebo arm.

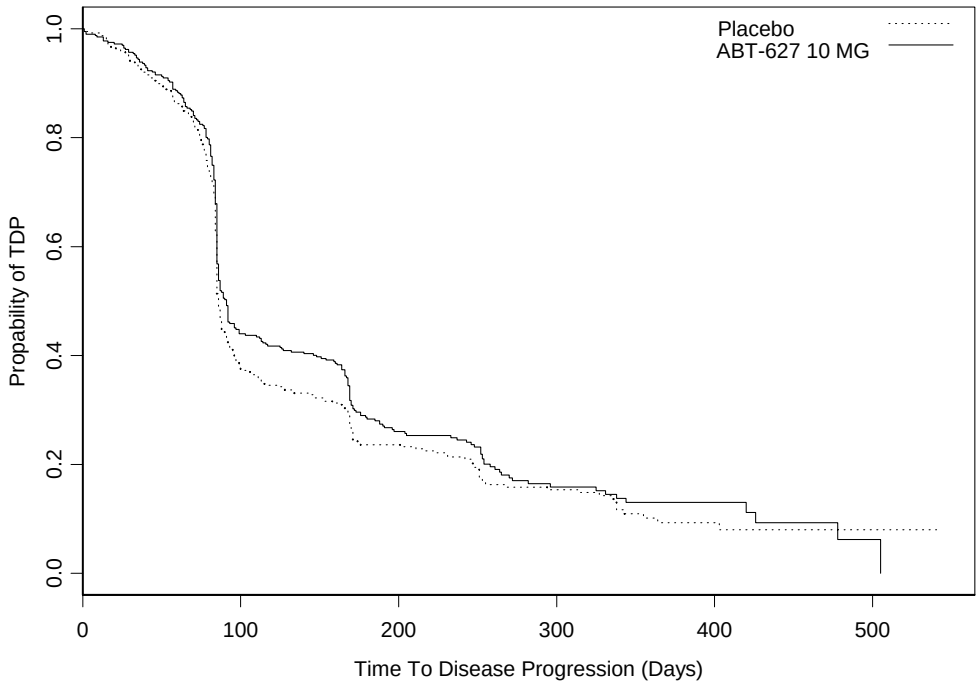

Figure 2: Kaplan-Meier Curves for TDP in the ITT Population

Reviewer's Comments:

- **The primary TDP analysis failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group (p = 0.143). This analysis used all of the two-sided alpha of 0.05.**

- Per the sponsor specified protocol:

"If the primary efficacy analysis is statistically significant at the α=0.05 level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule ……. If the primary efficacy analysis is not statistically significant at the α=0.05 level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values."

With the failed primary analysis, all pre-specified secondary and tertiary analyses are therefore considered as exploratory / hypothesis generating.

- The ICH E9 guidelines state that the statistical section of the protocol should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way in which anticipated analysis problems will be handled.

- Per-protocol analysis (Table 12) was considered as tertiary and the protocol clearly indicated that statistical significance will not be declared, regardless of the observed p-values. The sponsor after submission of NDA has presented the per-protocol analysis as one of the primary analyses. However, the protocol and the statistical plan were never amended to include per-protocol analysis (as primary efficacy analysis), or analyses on subgroup of patients with bone metastasis (Table 13). Furthermore, no statistical adjustment was made for the multiple analyses.

- According to the ICH E9 guidelines, in confirmatory analyses, adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan. The ICH E9 guidelines also state that an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol.

3.1.1.6.3 Secondary Efficacy Analyses

Mean change from baseline to final value in bone alkaline phosphatase

Three hundred sixty-four patients (89%) in the 10 mg Atrasentan arm had both baseline and final values in bone alkaline phosphatase and 374 patients (93%) in the placebo arm. The means change from baseline to final value in bone alkaline phosphatase measurement were: 13.19 in the treatment arm and 33.86 in the placebo arm. ANCOVA with treatment group and baseline bone alkaline phosphatase as the covariates obtained an unadjusted p-value of 0.001.

Reviewer's Comments:

According the protocol, if the primary efficacy analysis is not statistically significant at the α=0.05 level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values.
With the failed primary analysis, the p-value of 0.001 is not interpretable. For bone alkaline phosphatase, 11% of patients in the Atrasentan arm did not have both baseline and final measures; 7% in the placebo arm.

Time to Onset of PSA Progression

There were 236 events for PSA progression in the Placebo arm and 223 events in the 10 mg Atrasentan arm. Among the 236 events in the placebo arm, there were

81 events from USA. In the treatment arm, 67 events were from USA. There were 52 patients (12.7%) in the Atrasentan did not have time to PSA progression and 37 patients (9.2%) in the placebo arm. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the unadjusted p-value was 0.344. The hazard ratio for the time-to-PSA progression in the 10 mg Atrasentan arm, as compared with the Placebo arm, was 0.841.

The results from the stratified $G^{1,1}$ test and unstratified log-rank test are presented in the Table 5. The Kaplan-Meier curves for the ITT population are illustrated in Figure 3.

Table 5. Time To PSA Progression Analysis in ITT Population

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	356	364
Number of events (%)	223 (62.6%)	236 (64.8%)
Median (days), 95% CI	**87 (85, 162)**	**85 (71, 85)**
Stratified $G^{1,1}$ test	p=0.294	
Unstratified Logrank test	p=0.051	
Hazard ratio (95% CI)[1]	0.839 (0.698, 1.008)	

[1]: Hazard Ratio for PSA progression in the 10 mg Atrasentan arm, as compared with the Placebo arm.

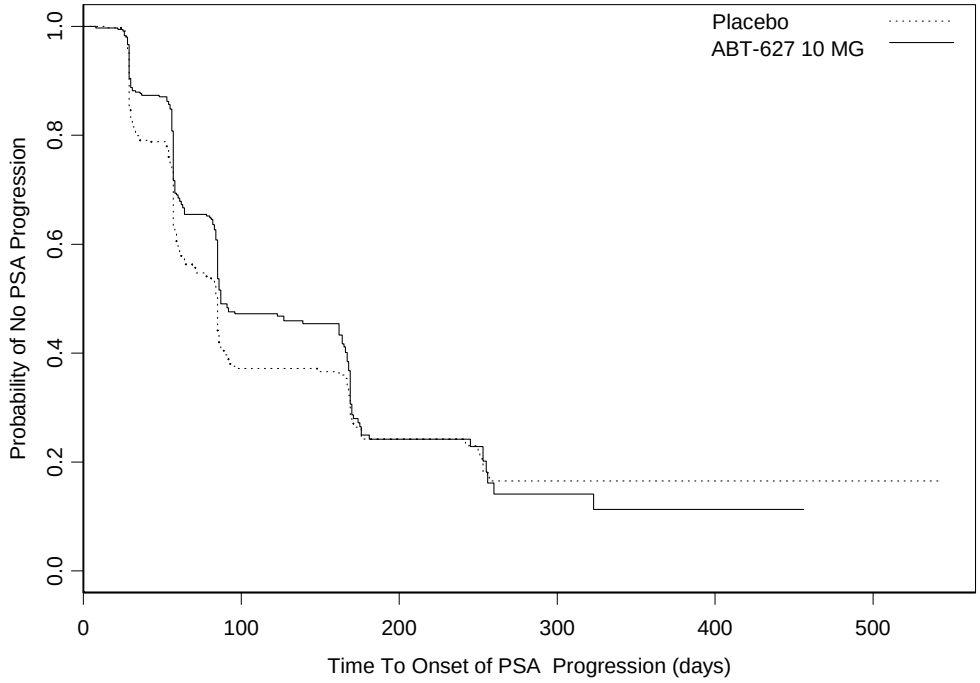

Figure 3: Kaplan-Meier Curves for Time to PSA Progression

Mean Rate of Change from Baseline to Final Value in Total Bone Scan Index

Two hundred ninety-four patients (72%) in the 10 mg Atrasentan arm had both baseline and final values in total bone scan index and 284 patients (71%) in the placebo arm. The means change from baseline to final value in total bone scan index was: 0.051 in the treatment arm and 0.053 in the placebo arm. ANCOVA with treatment group and total baseline bone scan index as the covariates obtained an unadjusted p-value of 0.844.

Overall Survival

Overall Survival (OS) analysis was performed using data submitted on February 24, 2005. There were 227 events for overall survival in the placebo arm and 234 events in the 10 mg Atrasentan arm. Among the 227 events in the placebo arm, there were 67 events from USA. In the treatment arm, 69 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the unadjusted p-value was 0.791. The hazard ratio for the OS in the 10 mg Atrasentan arm, as compared with the placebo arm, was 0.992.

The results from the stratified $G^{1,1}$ test and unstratified log-rank test are presented in the Table 6 (survival data was adjudicated by the medical reviewer). The Kaplan-Meier curves for the ITT population are illustrated in Figure 4.

Table 6. OS Analysis in ITT Population

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	408	401
Number of events (%)	234 (57.4%)	227 (56.6%)
Median (days), 95% CI	**574 (505, 638)**	**567 (525, 631)**
Stratified $G^{1,1}$ test	**p=0.791**	
Unstratified Logrank test	**p=0.9290**	
Hazard ratio (95% CI)[1]	**0.982 (0.818, 1.180)**	

[1]: Hazard Ratio for death in the 10 mg Atrasentan arm, as compared with the Placebo arm.

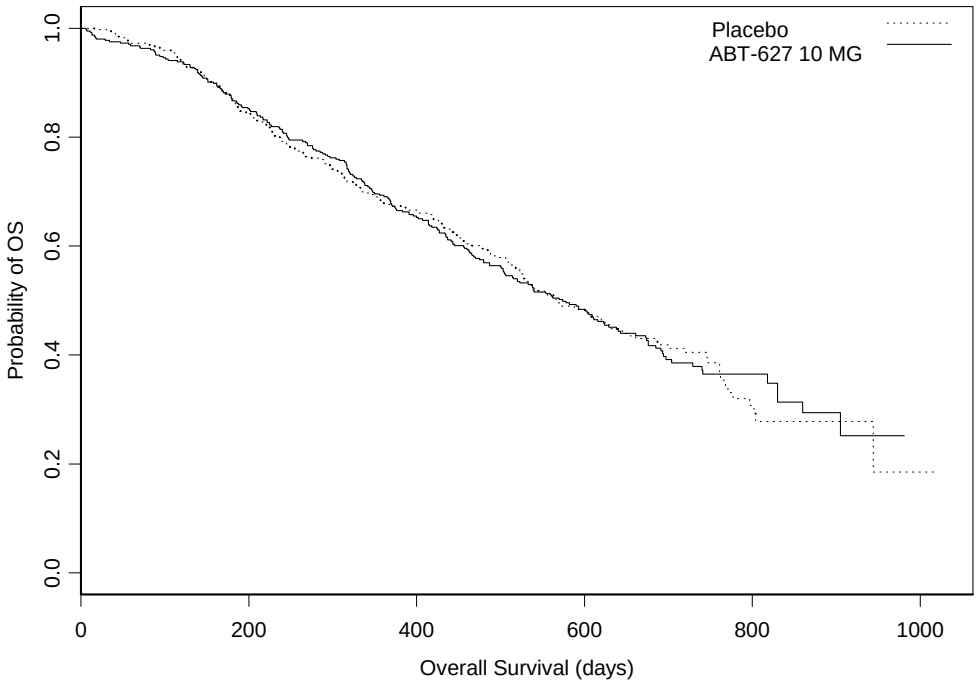

Figure 4: Kaplan-Meier Curves for OS in the ITT Population

Reviewer's Comments:

Per the sponsor specified protocol:

"If the primary efficacy analysis (time-to-disease progression using the stratified $G^{1,1}$ test) is statistically significant at the n=0.05 level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule"

"if any of these secondary analyses does not achieve statistical significance at the α=0.05 level, then statistical significance will not be declared for the subsequent secondary analyses, regardless of the observed p-values. If the primary efficacy analysis is not statistically significant at the α=0.05 level then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p values".

Because this study failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group (p = 0.143) and used all of the two-sided alpha of 0.05, all the four pre-

specified secondary endpoint analyses are considered as exploratory / hypothesis generating.

The results from the OS analysis suggest that Atrasentan was no better than placebo. The results using data from the medical reviewer were similar to the sponsor's.

3.1.1.6.4 Quality of life

Quality of life (QoL) was assessed using two scales: the Functional Assessment of Cancer Therapy – Prostate (FACT-P) and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). The mean change from baseline to final assessment was analyzed using ANCOVA. This review will focus on FACT-P assessment.

According the protocol, the FACT-P Version 4.0 instrument used in this study was tailored to explore issues specific to prostate cancer. It comprises the FACT-G general cancer scale with the addition of a 12-question prostate cancer–specific subscale (PCS) focusing on questions of pain and urinary and sexual function. The FACT-G is a 34-item generic QoL instrument, which measures four important dimensions: physical well-being, social/family well-being, emotional well-being, and functional well-being. The FACT-P composite score is the sum of the physical and functional well-being scores and the PCS. Each scale has a unique range; for all scales, results are totaled and calculated such that higher values represent better quality of life. Results for all domains were collected over the course of the study at baseline (Day 1), Week 4, at 12-week intervals thereafter, at the final visit, and at any safety follow-up visits.

The mean changes from baseline to the final time the subject completed the FACT-P QoL questionnaires are represented in Table 7 (the sponsor's Table 8 in the study report).

Table 7. Mean Change from Baseline to Final Assessment for FACT-P and Subscores: ITT Subject Population

	Mean Change from Baseline at Final Assessment		Treatment Difference		
Scale (Range)	Placebo	Atrasentan	Mean	SE	P-value
FACT-P Grand Total (0–156)	–9.34	–7.95	1.40	1.222	0.253
FACT-G Total (0–108)	–6.11	–5.54	0.58	0.874	0.510
Prostate Cancer Subscore (0–48)	–3.36	–2.34	1.02	0.476	0.032*
Composite (0–104)	–8.54	–7.76	0.78	0.986	0.430
Physical well-being (0–28)	–2.71	–2.73	–0.02	0.343	0.946
Social/family well-being (0–28)	–0.17	0.26	0.44	0.278	0.115
Emotional well-being (0–24)	–0.70	–0.49	0.21	0.265	0.438
Functional well-being (0–28)	–2.59	–2.57	0.02	0.358	0.956

*Represents a statistically significant difference between treatment groups at P = 0.05.
Cross Reference: Table 14.2__9.1.1

Reviewer's Comments:

QoL was defined as a tertiary analysis. The statistical plan was never amended to include QOL as primary efficacy analysis. No statistical adjustment was made for the multiple analyses.

Per the sponsor, *in the ITT analysis, Atrasentan treatment resulted in a difference in QoL in favor of Atrasentan as measured by the disease-specific PCS score (P = 0.032).* This review questioned this statement because of the following reasons:

1. The protocol did not pre-specify the statistical hypothesis, particularly the alternative hypothesis, which was used in the testing procedure. Hence, we do not know whether the PCS mean change is meaningful. Also with the PCS scores ranging from a possible 0 to 48, it is difficult to interpret the observed PCS mean change of 1.02.
2. Due to missing values, the PCS analysis did not include all patients. There were 32 patients in placebo arm who had missing values; 43 patients in the Atrasentan arm. Therefore, the analysis was not based on the ITT population.
3. Table 7 included eight tests. No statistical adjustment was made for the multiple analyses. None of them had p-value <0.05 except for the PCS score. The p-value of 0.032 for the PCS score is not interpretable.
4. Both QoL instruments ask the patients to rate the symptom based on their experience over the past week.

3.1.2 Study M96-594

Study M96-594 was a Phase 2, randomized, double-blind, placebo-controlled, parallel-group, multicenter study of Atrasentan. The primary objective of this dose ranging study was to assess the safety and efficacy of Atrasentan combined with supportive treatment, compared to placebo combined with supportive treatment, in clinically asymptomatic subjects diagnosed with hormone regractory prostate cancer (HRPCa).

This study consisted of a 14-day screening period and a double-blind treatment period that continued unti1 the subject experienced clinica1 disease progression or otherwise discontinued from the study. At least 204 subjects were to be enrolled to ensure 68 subjects per treatment arm. Subjects were randomized to receive: 10 mg Atrasentan, 2.5 mg Atrasentan, or placebo daily in a 1:1:1 ratio, in addition to their standard care. A subject was considered to have completed the study after presenting evidence of disease progression (event).

Design
The study consisted of a 14-day screening period and a double-blind treatment period, which continued until the subject either experienced clinical disease progression or otherwise discontinued from the study (Figure 5). Subjects who experienced disease progression were eligible to enter an open-label extension study, M97-739. Subjects visited the site on day 1 and at 2, 4, 6, 8, 10, 12, 18, 20, 22, and 24 weeks, and every 4 weeks thereafter until the final visit. Follow-up assessments were performed 28 days later; there was also a 2-year survival follow-up visit. Tumor markers were measured at baseline and every scheduled visit. Bone markers were measured at baseline and at 6, 12, and 24 weeks, and every 4 weeks thereafter until final visit. QOL questionnaires were completed at baseline and at 6, 12, and 24 weeks, and every 4 weeks thereafter until final visit. Performance status was evaluated at baseline, 2, 6, 12, and 24 weeks, and every 4 weeks thereafter until final visit. Bone and computed tomography (CT) scans were performed at baseline, at the final visit, and at the investigator's discretion. After the blind break on 31 January 2000, active subjects who had been randomized to receive 10 mg Atrasentan were allowed to continue receiving open-label Atrasentan and those who had been randomized to receive either 2.5 mg Atrasentan or placebo were allowed to receive 2.5 mg or 10 mg Atrasentan open-label.

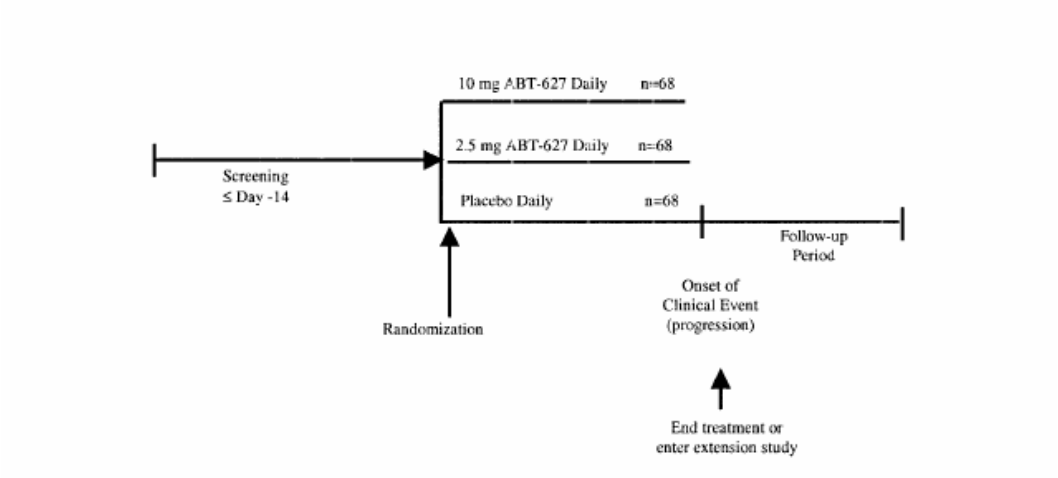

Figure 5. Study Design Schematic (M96-594)

A total of 288 men with asymptomatic, metastatic HRPC were enrolled in study M96-594: 104 in the placebo arm, 95 in the 2.5 mg Atrasentan arm, and 89 in the 10 mg Atrasentan arm.

Objectives and efficacy endpoints:

The primary objective of this dose ranging study was to assess the safety and efficacy of Atrasentan combined with supportive treatment, compared to placebo combined with supportive treatment, in clinically asymptomatic subjects diagnosed with HRPCa.

The primary endpoint was time to disease progression (TDP). It was defined as the time interval from the time of randomization, defined as the first dose of study drug, to the first onset of **an investigator-determined** occurrence of any of the following events:

1. Palliative treatment of new bone or visceral pain with an opioid
2. Palliative radiation
3. Treatment with chemotherapy or increased steroid use (glucocorticoids)
4. New tumor growth–related symptoms
5. New measurable bone lesions
6. New measurable soft-tissue lesions
7. Other investigator-defined measures of disease progression

Overall surviva1, time-to-PSA progression, bone biology markers, biochemical markers of tumor burden, bone scan indices, quality of life, and quality adjusted time- to-progression results were secondary efficacy variables. All subjects were followed for survival, defined as the time from receipt of first dose of study drug until death. Survival data for subjects still alive at the time of analysis or lost to

follow-up were censored at the date of last contact. PSA va1ues were measured and the time-to- event of an increase of serum PSA of > 50% from baseline on two occasions at least 2 weeks apart was analyzed.

Reviewer's Comments:

M96-594 used a different time-to-disease-progression (TDP) definition from the study M00-211 (Table 8). Therefore, its TDP results of M96-594 cannot be used as supportive evidence to the study M00-211.

Table 8. Comparison of TDP Definitions in M00-211 and M96-594

M00-211	M96-594
TDP was measured **from randomization to the first occurrence** of any of the following events: 1. Pain due to prostate cancer requiring use of opioids, glucocorticoids, radionuclides, radiation, or chemotherapy 2. A skeletal- related event, ie, a pathologic or vertebral compression fracture not related to trauma 3. A metastatic prostate cancer–related event requiring intervention 4. A bone scan after baseline demonstrating 2 or more new skeletal lesions read by an independent radiologist 5. A CT or MRI scan subsequent to baseline demonstrating evidence of extra- skeletal disease progression read by an independent radiologist	TDP was measured **from the first dose of study drug to the first of an *investigator-determined* occurrence** of any of the following events: 1. Palliative treatment of new bone or visceral pain with an opioid 2. Palliative radiation 3. Treatment with chemotherapy or increased glucocorticoids use 4. New tumor growth-related symptoms 5. New measurable bone lesions 6. New measurable soft-tissue lesions 7. Other investigator-defined measures of disease progression

Sample Size Considerations

The total number of subjects initially planned for this study was based on the assumptions that the accrual and duration of the study would be 6 and 18 months, respectively, and that the median time to disease progression in the placebo treatment group would be 6 to 9 months. Under this assumption, 204 subjects, 68 per arm, would be sufficient to detect a 50% improvement of time to disease progression in one of the Atrasentan dose groups compared to the placebo group with 0.2 significance and 70% power. The total number of subjects with disease progression in the 204 subjects during the 18-month study duration was projected to be 125.

Enrollment was surpassed by 84 subjects and the total number of subjects with disease progression in the 288 subjects was 202 at the time of the primary analysis cutoff date (31 October 1999).

Efficacy Analysis Methods

The primary efficacy measurement was time-to-progression. The distribution of time to progression was estimated using Kaplan-Meier methodology. Comparison between randomization groups in time to progression was performed primarily using a log-rank test. All data were analyzed according to intent-to-treat principle.

Primary Efficacy Analysis

Medians time to disease progression for the three study groups were: 137 days in the placebo arm, 178 days in the 2.5 mg Atrasentan arm, and 183 days in the 10 mg Atrasentan arm. This review will focus on the comparison of 10 mg Atrasentan arm and the placebo ram without multiple comparison adjustment. Unstratified log-rank test showed that p-value for the 10 mg Atrasentan arm vs. Placebo was 0.1323 (Table 9). The Kaplan-Meier curves for the ITT population are illustrated in Figure 6.

Table 9. Primary Efficacy TDP Analysis in ITT Population (M96-594)

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	89	104
Number of events (%)	58 (65.2%)	77 (74.0%)
Median (days), 95% CI	183 (132, 225)	137 (116, 167)
Unstratified Logrank test	**p=0.1323**	
Hazard ratio (95% CI)[1]	0.769 (0.545, 1.085)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the placebo arm.

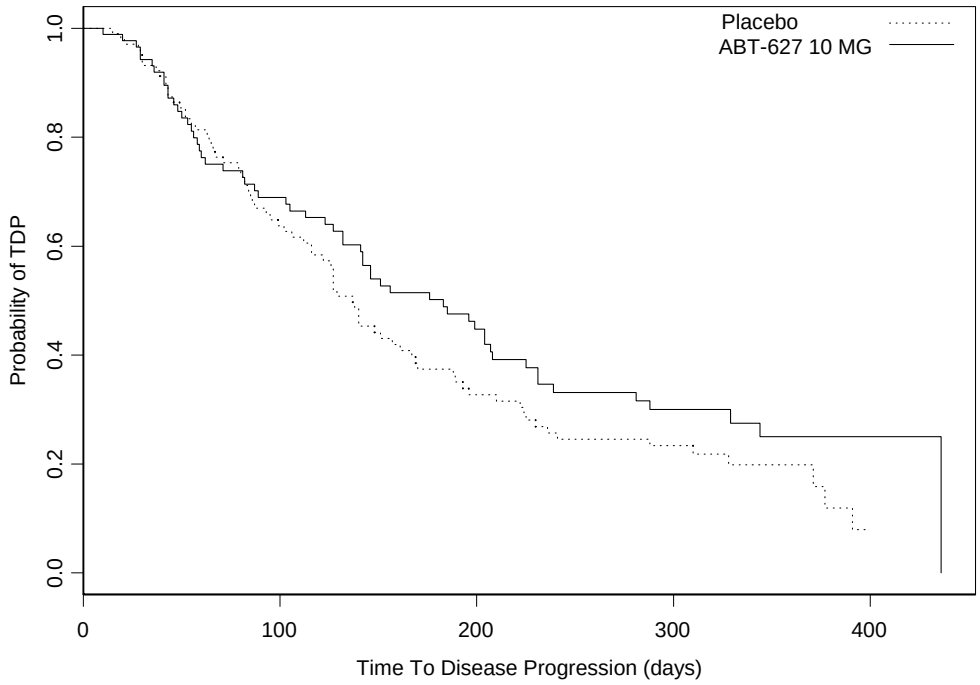

Figure 6: Kaplan-Meier Curves for TDP in ITT Population (M96-594)

Secondary Efficacy Analysis

Analysis of survival is complicated by the crossover of placebo-treated subjects to open-label Atrasentan treatment. At the time of disease progression in study M96-594 all subjects were eligible to participate in study M97-739, an open-label extension study in which subjects were randomized to receive either 20 or 30 mg Atrasentan. In addition, subjects who remained in Study M96-594 when the blind was broken (31 January 2000) were eligible to receive open-label drug (2.5 mg or 10 mg) and continue in the study until disease progression occurred. Overall, approximally 59% of Atrasentan subjects in study M96-594 continued to receive Atrasentan after disease progression or blind break, while 57% of subjects treated with placebo ultimately received Atrasentan therapy.

Final database closure occurred on 30 June 2002. As of this date, the number of confirmed deaths were 70/104 (67%), 61/95 (64%), and 57/89 (64%) for the placebo, 2.5 mg Atrasentan, and 10 mg Atrasentan groups, respectively. **There were no statistically significant differences in survival** between either of the Atrasentan treatment group and the placebo group (Table 10). The Kaplan-Meier curves for the ITT population are illustrated in Figure 7.

Table 10. OS Analysis in ITT Population (M96-594)

	10 mg Atrasentan	Placebo
Number of patients (ITT)	89	104
Number of events (%)	57 (64.0%)	71 (68.3%)
Median (days), 95% CI	**528 (490, 675)**	**534 (433, 632)**
Unstratified Logrank test	**p=0.5905**	
Hazard ratio (95% CI)[1]	**0.909 (0.641, 1.289)**	

[1]: Hazard Ratio for death in the 10 mg Atrasentan arm, as compared with the Placebo arm.

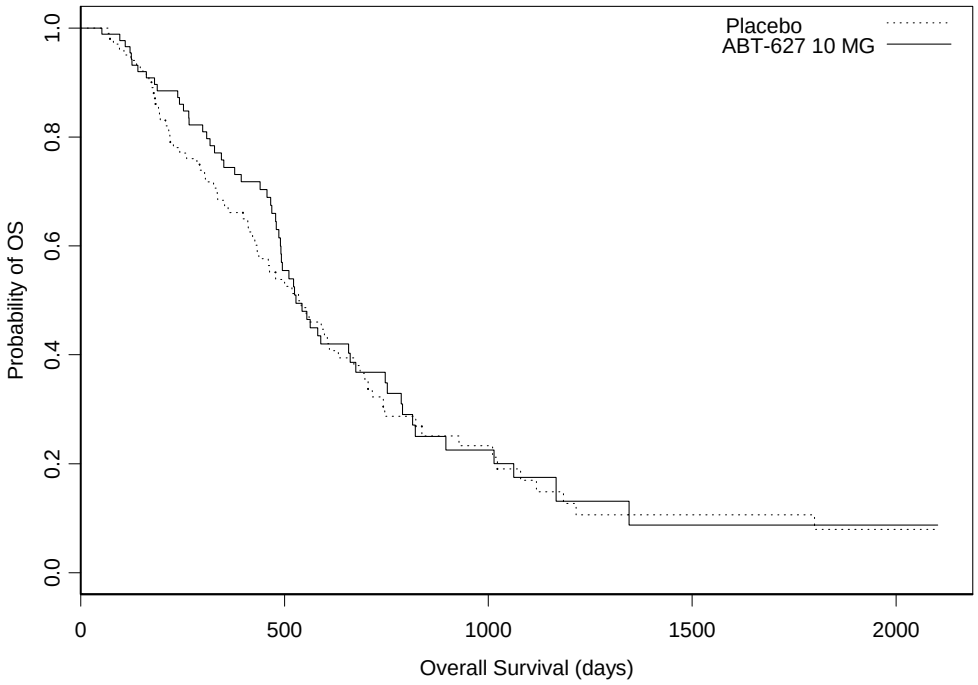

Figure 7: Kaplan-Meier Curves for OS in ITT Population (M96-594)

Reviewer's Comments:

Study M96-594 is a dose ranging study, which included three study arms: placebo, 2.5 mg Atrasentan, and 10 mg Atrasentan. The design is different from the design used in the phase III study M00-211. This study failed to demonstrate efficacy based on time to disease progression and overall survival (no difference compared to placebo).

M96-594 used a different time-to-disease-progression (TDP) definition from the study M00-211. Therefore, its TDP results cannot be used as supportive evidence to the study M00-211.

3.2 Evaluation of Safety

Please refer to Clinical Review of this application for safety evaluation.

4 Findings in Special/Subgroup Populations in M00-211

4.1 Gender, Race and Age

For each subgroup population, stratified $G^{1,1}$ test and unadjusted log-rank test were performed. Because all the patients studied in this application were men, no gender analysis will be conducted. The 95.2% of patients were Caucasian (15 patients in the placebo arm were not Caucasian; 24 patients in the treatment arm). No race subgroup analysis will be performed in this review. Therefore, this section will focus on TTP analyses by age (< 65 years vs. ≥ 65 years).

For patients with age ≥65, there were 256 events (78%) for disease progression in the Placebo arm and 246 events (72.7%) in the 10 mg Atrasentan arm. Among the 256 events in the placebo arm, there were 84 events from USA. In the treatment arm, 87 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the unadjusted p-value was 0.140. The hazard ratio for the time-to-disease progression in the 10 mg Atrasentan arm, as compared with the placebo arm, was 0.881.

For patients with age <65, there were 59 events (75.6) for disease progression in the Placebo arm and 53 events (77.9%) in the 10 mg Atrasentan arm. Among the 59 events in the placebo arm, there were 14 events from USA. In the treatment arm, 13 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the unadjusted p-value was 0.693. The hazard ratio for the time-to-disease progression in the 10 mg Atrasentan arm, as compared with the placebo arm, was 0.858.

The results from the stratified $G^{1,1}$ test and unstratified log-rank test are presented in the Table 11.

Table 11. TDP Analyses by Age in ITT Population
(Base on the data submitted on December 19, 2003)

	10 mg Atrasentan	Placebo
Age		
<65		
Number of patients (ITT)	68	78
Number of events (%)	53 (77.9%)	59 (75.6%)
Median (days), 95% CI	**87 (85, 151)**	**85 (84, 91)**
Stratified $G^{1,1}$ test	p=0.693	
Unstratified Logrank test	**p=0.582**	
Hazard ratio (95% CI)[1]	0.858 (0.587, 1.253)	
>=65		
Number of patients (ITT)	340	323
Number of events (%)	246 (72.3%)	252 (78.0%)
Median (days), 95% CI	**91 (86, 99)**	**87 (85, 92)**
Stratified $G^{1,1}$ test	p=0.140	
Unstratified Logrank test	**p=0.154**	
Hazard ratio (95% CI)[1]	0.881 (0.739, 1.051)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the Placebo arm.

Reviewer's Comments:

Eighty-two percent of patients in this study were more than 65 year old. For those patients, the hazard ratio was similar to the hazard ratio observed in the entire patient population of the study. There appears to be little or no effect in both age groups.

4.2 Other Special/Subgroup Populations

The sponsor submitted two main subgroup analyses: per-protocol analysis and analysis for patients with bone metastases at baseline. All the analyses presented in this section are considered exploratory/ hypothesis generating.

Per − Protocol Subject Population (M00-211)

The per-protocol analysis excluded the following patients: 1) men without definitive evidence of metastatic HRPCa, 2) men with insufficient evidence of a hormone-refractory state, and 3) other factors including minimal study drug exposure and potentially confounding medication, e.g., opiates or not having received study drug. Among 671 patients were included in this analysis: 342 patients in the Atrasentan arm and 329 patients in the Placebo arm.

The datasets (tte01_ar.xpt or tte02_ar.xpt) had 670 subjects with PERPTRCL = 'YES', but there were a total of N=671 subjects in the analyses of the per protocol subset. Per the sponsor, during the classification process a total of 139 subjects

were identified as subjects with data excluded from the ITT population. However, there is one subject (ptno= 1312) that received exclusionary medications on 22SEP02 and was censored after that day, i.e. this subject is still included in the per-protocol analyses up to that date.

There were 267 events (81.2%) for disease progression in the Placebo arm and 255 events (74.6%) in the 10 mg Atrasentan arm. Among the 267 events in the placebo arm, there were 84 events from USA. In the treatment arm, 81 events were from USA. A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the unadjusted p-value was 0.018. The hazard ratio for the time-to-disease progression in the 10 mg Atrasentan arm, as compared with the placebo arm, was 0.804 (Table 12). The Kaplan-Meier curves for the per-protocol population are illustrated in Figure 8.

Table 12. TDP Analysis in Per-Protocol Population (M00-211)

	10 mg Atrasentan	**Placebo**
Number of patients (ITT)	342	329
Number of events (%)	255 (74.6%)	267 (81.2%)
Median (days), 95% CI	90 (86, 97)	85 (85, 88)
Stratified $G^{1,1}$ test*	p=0.018*	
Unstratified Logrank test*	p=0.011*	
Hazard ratio (95% CI)[1]	0.804 (0.676, 0.956)	

[1]: Hazard Ratio for progression in the 10 mg Atrasentan arm, as compared with the placebo arm.
* unadjusted for multiplicity

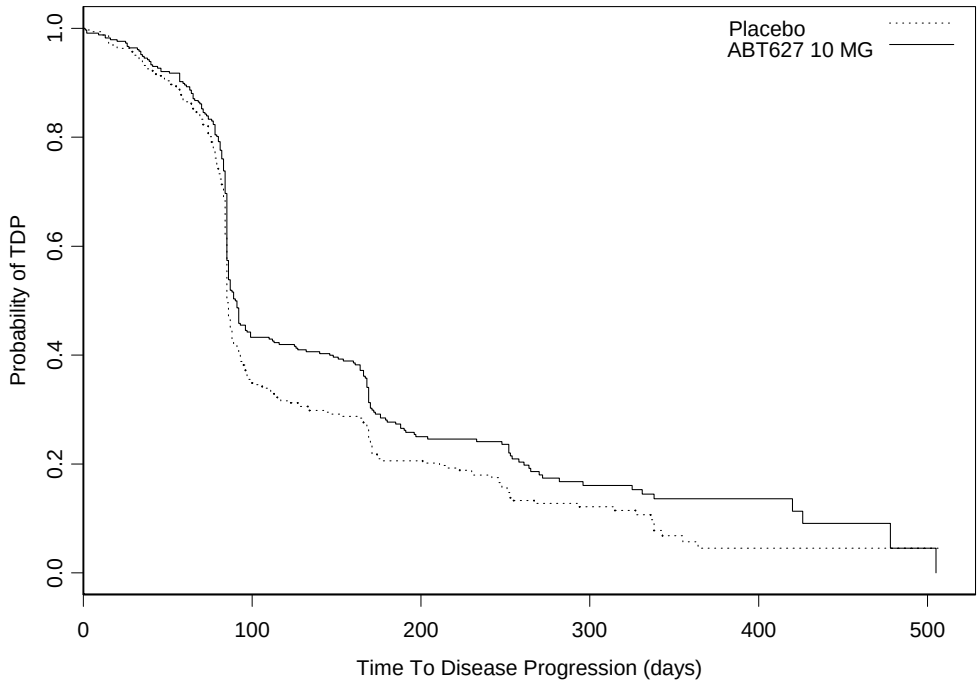

Figure 8: Kaplan-Meier Curves for TDP in the Per Protocol Population

Patients with Bone Metastases at Baseline (M00-211)

Per the sponsor, in the M00-211, there are 684 subjects presenting with bone metastases at baseline, where the baseline scans were performed prior to study drug administration. Subsequently, the definition of baseline was changed to include bone and CT scans up to 21 days after the start of study drug to account for subjects whose baseline scans were performed after the start of study drug. As a result, six additional patients (1038, 1039, 1354, 2693, 1013, and 2899) were added to this cohort for a total of 690 patients with bone metastases at baseline. FDA analysis did not include these 6 patients (Table 13). A stratified (US sites vs. non-US sites) $G^{1,1}$ test showed that the p-value was 0.022 (unadjusted for multiplicity). The hazard ratio for the time-to-disease progression in the 10 mg Atrasentan arm, as compared with the Placebo arm, was 0.807 (Table 13). The Kaplan-Meier curves for the patients with bone metastases at baseline are illustrated in Figure 9.

Table 13. TDP Analysis in Patients with Bone Metastases at Baseline (FDA Analysis)

	10 mg Atrasentan	Placebo
Number of patients (ITT)	352	332
Number of events (%)	259 (73.6%)	262 (78.9 %)
Median (days), 95% CI	92 (86, 99)	85 (85, 87)
Stratified $G^{1,1}$ test*	p=0.020*	
Unstratified Logrank test*	p=0.0114*	
Hazard ratio (95% CI)[1]	0.798 (0.671, 0.950)	

[1]: Hazard Ratio for recurrence or death in the 10 mg Atrasentan arm, as compared with the placebo arm; * unadjusted for multiplicity

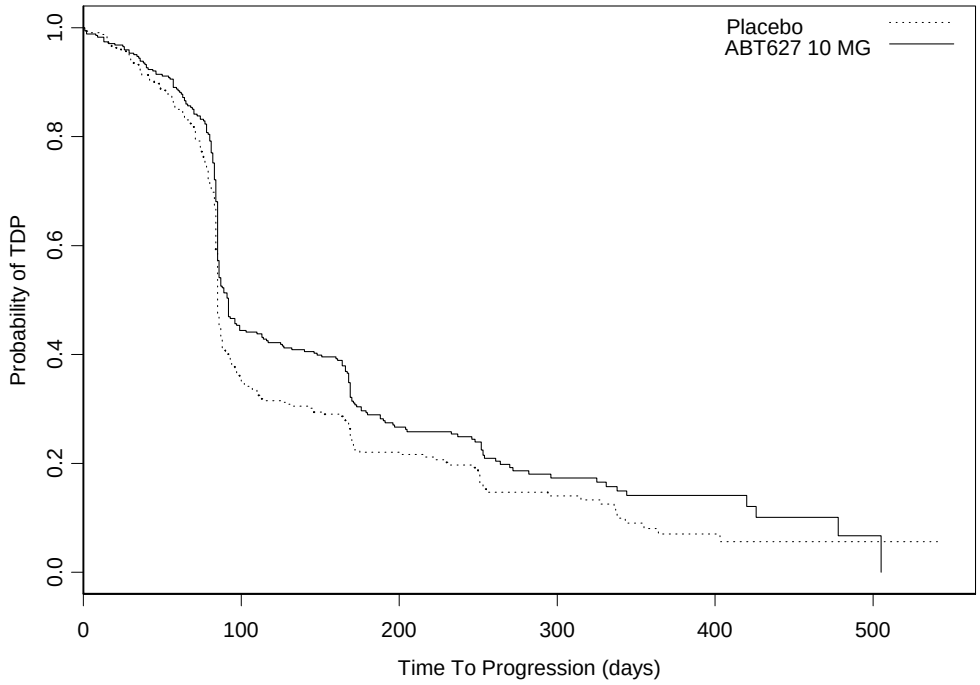

Figure 9: Kaplan-Meier Curves for TDP in Patients with Bone Metastases at Baseline

Other Subgroup Analyses (M00-211)

The sponsor also did TDP analyses on the following subpopulations, which were defined retrospectively:

- Patients with no bone metastases at baseline
- Patients with soft-tissue metastases at baseline
- Patients with no soft-tissue metastases at baseline
- Patients with both bone and soft-tissue metastases at baseline

- Patients with bone but no soft-tissue metastases at baseline
- Patients with soft-tissue but no bone metastases at baseline
- Patients with no metastases at baseline

None of the above subgroup analyses had p-value <0.05 except for the patients with bone but no soft-tissue metastases at baseline and the patients with no metastases at baseline. With the failed primary analysis and no pre-specification of these analyses, the p-values from these analyses are not interpretable.

Exploratory Analyses

In order to evaluate if the time of bone scan influenced the primary outcome the following exploratory analyses were conducted.

Time from randomization to each bone scan was calculated. Means and standard deviations of bone scan times are presented in Table 14.

Table 14. Mean and SD (in weeks) of Time To Bone Scan From Randomization

Time from randomization to Bone Scan	# (%)		Mean (SD)	
	Atrasentan N= 408	**Placebo N=401**	**Atrasentan N= 408**	**Placebo N=401**
Week 12	327	328	11.9 (1.3)	11.9 (1.4)
Week 24	135	107	24.2 (1.3)	23.8 (0.9)
Week 36	49	54	35.6 (1.9)	36.0 (1.7)
Week 48	21	20	47.8 (0.8)	47.7 (1.0)
Week 60	6	7	60.4 (0.8)	59.5 (2.5)
Week 72	2	1	70.4 (2.3)	71.4 (-)

Log-rank test was used to test if cumulative percentages (survival curves) were equal. Results from the tests are presented in Table 15.

Table 15. Median (in Weeks) of Time to Bone Scan and Log-rank Test

Time from randomization to Bone Scan	**Atrasentan N= 408**	**Placebo N=401**	**Log-rank Test**
Week 12	12.0	12.0	0.8184
Week 24	24.0	24.0	0.0154
Week 36	35.9	36.0	0.1356
Week 48	47.9	47.7	0.8001
Week 60	60.4	60.0	0.7494
Week 72	70.4	71.4	-

The log-rank test showed that there was no difference between two distributions of time to assessment, except time to week 24 bone scan. However, two medians at time to week 24 bone scan were the same and the numbers of missing values were 273 (67%) in the Atrasentan arm and 294 (73%) in the placebo arm. The variation observed in the median TDP in the ITT and all the subgroups analyzed is within the variation (SD of 1-2 weeks) in the bone scan assessment.

Reviewer's Comments:

After the studies failed to demonstrate the efficacy, and after the submission of NDA to the Agency, the applicant is seeking approval based on a post-hoc, subgroup, exploratory analysis in the subgroup of patients with metastasis to bone and claims a favorable effect in this subpopulation (per the applicant, HR=0.813, CI= 0.685-0.965; p-value= 0.016 (unadjusted)).

The study M00-211 was designed to answer a question about the overall Atrasentan effect in the entire population, not to answer questions about the subgroups. The statistical plan was never amended to include per-protocol analysis and baseline bone metastasis subgroup analysis as primary efficacy analyses. Furthermore, the analysis for patients with bone metastases at baseline was not pre-specified. The observed TDP difference in the subgroup of patients with bone metastases at baseline, is considered as an exploratory and hypothesis generating analysis. The p-value from this subgroup analysis is not interpretable since all the type I error rate has been spent in the failed primary protocol pre-specified analysis.

Although a TDP difference was observed in the per-protocol patient population and in patients with bone metastases at baseline, these analyses were not pre-specified and this data has been analyzed multiple times and within multiple subgroups with no type I error adjustment. Hence, it can not be ruled out that these are false positive results. Without any type I error adjustment for the multiple comparisons, these analyses are only considered explorative/hypothesis generating. The protocol further indicated that significance will not be declared, regardless of the observed p-values. The findings from the subgroup analyses should be confirmed through other studies.

According to the ICH E9 guidelines, any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted. The ICH E9 also states that an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol. **Subgroup analyses usually have high false positive or false negative rate. Statistical results should be interpreted with extreme caution because false positive findings may increase as the number of significance tests performed increases.**

It is impossible to correctly adjust the nominal p-value for multiple comparisons post hoc. Generally, post hoc analyses are hypothesis generating. Subgroup analyses suggest hypotheses worth examining in other studies. The strength of evidence for efficacy is discredited with multiple subgroup analyses where one has many chances to find a difference between two arms, and can inflate the Type I error.

The primary ITT analysis of time-to-disease progression was pre-specified in the M00-211. However, this study did not pre-specify any adjustment procedure for the subgroup analyses using the per-protocol population or baseline bone metastatic patient population. According to the ICH E9, adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan. Without any allocation of type I error for the analyses using the per-protocol population or baseline bone metastatic patient population, the results from the analyses will be considered as supportive to the ITT analysis.

Pooled Analysis for study M00-211 and M96-594

This pooled analysis conducted by the sponsor included 1002 subjects randomized to receive either 10 mg Atrasentan or placebo in studies M00-211 and M96-594. It was performed using an integrated data set of subject-specific data from both studies. A log-rank analysis stratified by study was used as the primary integrated analysis for all time-to-event efficacy variables including time to disease progression

TDP Results from the pooled Analysis

According the study report, the integrated log-rank analysis of the intent-to-treat population from studies M00-211 and M96-594 with stratification by study demonstrates that 10 mg Atrasentan significantly delayed time to disease progression in men with metastatic HRPC (treatment difference in median time to disease progression was 46 days in study M96-594 and 5 days in study M00-211; log-rank $P = .045$). Similar results were obtained from the pooled $G^{1,I}$ analysis stratified by study ($P = .035$).

Reviewer's Comments:

M96-594 was a phase 2 study which randomized 3 arms (placebo, 2.5mg and 10mg of Astrasentan); M00-211 was a phase 3 study which had two arms (10mg Astrasentan and placebo). The pooled TDP analysis is not acceptable and would not support the effectiveness of Astrasentan because of the following reasons:

1. Both studies had different designs;
2. Both studies had different TDP definitions (Table 8);

3. No statistical tests for heterogeneity between two studies have been conducted;
4. Both studies were negative. (Please see Points To Consider On Application With Meta-Analyses by Committee for Proprietary Medicinal Products, EMEA, May 31, 2001);
5. In general when considering pooled analysis, a fixed-effects model or a random-effects model should be used to estimates the overall treatment effect;
6. No independent review of progression evaluation was conducted in study M96-594;
7. This was not a pre-specified analysis.
8. For the pooled analysis, it not clear how type I error is controlled.

Efficacy demonstration should be solely based on results of the primary analysis from individual trials, where the primary analysis is pre-specified and agreed upon by the Agency. Efficacy claims based on results from any other analyses (such as pre-specified exploratory or post-hoc) can only inflate the false positive error rate and may not be considered for regulatory approval.

5 Summary and Conclusions

By submitting this NDA application, the sponsor is seeking approval of using Xinlay (trade name Atrasentan, 10mg, orally administered) in the treatment of male subjects diagnosed with metastatic, hormone-rcfractory prostate cancer.

The sponsor has submitted results of the final analysis from the M00-211 and M96-594 studies designed to evaluate the efficacy of 10 mg Atrasentan for the treatment of metastatic HRPC. M00-211 was a phase 3, randomized, double-blind, placebo-controlled, multicenter, multinational study of 10 mg Atrasentan. Study M96-594 was a Phase 2, randomized, double-blind, placebo-controlled, parallel-group, multicenter study of Atrasentan.

Both studies M00-211 and M96-594 failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group.

5.1 Statistical Issues and Collective Evidence

1. The independent data monitoring committee (IDMC) recommended closure of study enrollment on 27 September 2002, when 809 subjects were enrolled and randomized. The IDMC determined that the null hypothesis was not likely to be rejected for the primary endpoint using the $G^{1,1}$ analysis on the intent-to-treat population. The decision was based on 809 subjects, 343 of which experienced disease progression**.** The last subject's last dose of administration of the study drug was on 19 March 2003.

2**.** The primary TDP analysis in M00-211 failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group ($p = 0.143$). This analysis used all of the two-sided alpha of 0.05.

   Per the sponsor specified protocol:
   "If the primary efficacy analysis is statistically significant at the $\alpha=0.05$ level, then p-values for the secondary analyses will be subject to multiple comparison adjustments using the step-down rule ……. If the primary efficacy analysis is not statistically significant at the $\alpha=0.05$ level, then statistical significance will not be declared for any of these secondary analyses, regardless of the observed p-values."

   With the failed primary analysis, all pre-specified secondary and tertiary analyses were considered as exploratory / hypothesis generating.

3. Study M00-211 was designed to answer a question about the overall Atrasentan effect in the entire population, not to answer questions about the subgroups. The protocol or the statistical plan was never amended to include per-protocol analysis and baseline bone metastasis sub-group analysis as primary efficacy analyses. Furthermore, the analysis for patients with bone metastases at baseline was not pre-specified. Although the per-protocol analysis was outlined in the protocol, it was considered as tertiary. The protocol further indicated that significance for the subgroup analyses will not be declared, regardless of the observed p-values.

4. No statistical adjustment was made for the multiple analyses (subgroup analyses) and multiple hypotheses . Without any type I error adjustment for the multiple comparisons, these analyses are not interpretable. The findings from the analyses should be confirmed through other studies.

5. According to the ICH E9 guidelines (Statistical Principles for Clinical Trials), any conclusion of treatment efficacy (or lack thereof) or safety based solely on exploratory subgroup analyses is unlikely to be accepted. The ICH E9 also states that "…., an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol".

   Subgroup analyses usually have high false positive or false negative rate. Statistical results should be interpreted with extreme caution because false positive findings may increase as the number of significance tests performed increases.

6. It is impossible to correctly adjust the nominal p-value for multiple comparisons post hoc. Post hoc analyses are considered as hypothesis generating. Subgroup analyses suggest hypotheses worth examining in other studies. The strength of evidence for efficacy is discredited with multiple subgroup analyses where one could have many chances to find a difference between two arms. With no pre-specified analyses, the Type I error will also be inflated.

7. The primary ITT analysis of time-to-disease progression was pre-specified in the M002-11. However, the M00211 did not pre-specify any adjustment procedure for the subgroup analyses using the per-protocol population or baseline bone metastatic patient population. According to the ICH E9, adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan. Without any allocation of type I error for the analyses using the per-protocol population or baseline bone metastatic

patient populations, the results from the subgroup analyses will be considered as supportive to the ITT analysis and may not be claimed in the label.

8. M96-594 is a smaller dose ranging study, which included three study arms: placebo, 2.5 mg Atrasentan, and 10 mg Atrasentan. The design is different from the design used in the phase 3 study M00-211.

9. M96-594 used a different time-to-disease-progression (TDP) definition from the study M00-211 (Table 8). No independent review of progression evaluation was conducted in study M96-594. The primary TDP analysis in this study also failed to demonstrate a delay in disease progression. Therefore, its TDP results cannot be used as supportive evidence to the study M00-211.

10. M96-594 was a phase 2 study which randomized 3 arms (placebo, 2.5mg and 10mg of Astrasentan); M000-211 was a phase 3 study which had two arms (10mg Astrasentan and placebo). This reviewer does not believe in the demonstration of efficacy based on results from pooling trials together, especially when (a) neither of the trials individually showed a statistically significant difference; (b) both studies had different definitions of TDP and no independent review of progression evaluation was conducted in study M96-594; (c) the proposed analysis for pooling trials together is a post-hoc analysis; (d) for the pooled analysis, it is not clear how type I error is controlled.

11. Efficacy demonstration should be solely based on results of the primary analysis from individual trials, where the primary analysis is pre-specified and agreed upon by the Agency. Efficacy claims based on results from any other analyses (such as pre-specified exploratory or post-hoc) can only inflate the false positive error rate and may not be considered for regulatory approval.

12. Quality of Life (QoL) was defined as a tertiary analysis in M00-211. QoL was assessed using two scales: the Functional Assessment of Cancer Therapy – Prostate (FACT-P) and the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30). The statistical plan was never amended to include QOL as primary efficacy analysis. No statistical adjustment was made for the multiple QoL analyses.

    The sponsor's study report stated that *in the ITT analysis, Atrasentan treatment resulted in a difference in QoL in favor of Atrasentan as measured by the disease-specific PCS score (P = 0.032).* This reviewer questions this statement because of the following reasons:

    1) The protocol did not pre-specify the statistical hypothesis, particularly the alternative hypothesis, which was used in the testing procedure. Hence,

we do not know whether the PCS mean change is meaningful. Also with the PCS scores ranging from a possible 0 to 48, it is difficult to interpret the observed PCS mean change of 1.02.

2) Due to missing values, the PCS analysis did not include all patients. There were 32 patients in placebo arm who had missing values; 43 patients in the Atrasentan arm. Therefore, the analysis was not based on the ITT population.
3) Several statistical tests were performed without any statistical adjustment for the multiple analyses. The p-value of 0.032 is not interpretable.
4) Both QoL instruments ask the patients to rate the symptom based on their experience over the past week.

5.2 Conclusions and Recommendations

In this reviewer's opinion this NDA failed to demonstrate a delay in disease progression for subjects in the Atrasentan treatment group compared with those in the placebo group and the observed data does not support the sponsor's claim of efficacy of Atrasentan.

SIGNATURES/DISTRIBUTION LIST

Primary Statistical Reviewer: Shenghui Tang, Ph.D.
Date:

Concurring Reviewer: Rajeshwari Sridhara, Ph.D.
Team Leader

Kooros Mahjoob, Ph.D.,
Acting Director, Division of Biometrics
Date:

cc:
HFD-150/C. Cottrell
HFD-150/J. Johnson
HFD-150/A. Ibrahim
HFD-710/R. Sridhara
HFD-710/K. Mahjoob
HFD-700/C. Anello

C:\AAA\NDA\2005\Atrasentan\Reports\Xinlay(Atrasentan) Stat Review.doc