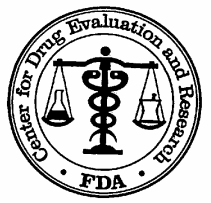


DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA #/Serial #:	20-727
DRUG NAME:	Bidil (hydralazine HC1 and isosorbide dinitrate)
INDICATION:	Treatment of heart failure in black patients
APPLICANT:	NitroMed, Inc.
DATE:	December 23, 2004
DATE OF COMPLETION:	May 4, 2005
REVIEW PRIORITY:	P
BIOMETRICS DIVISION:	Division of Biometrics I
STATISTICAL REVIEWER:	H.M. James Hung, Ph.D. (HFD-710)
MEDICAL DIVISION:	Division of Cardio-Renal Drug Product (HFD-110)
CLINICAL TEAM:	Salma Lemtouni, M.D. (HFD-110)
PROJECT MANAGER:	Dianne Paraoan (HFD-110)

KEY WORDS: composite score of death, heart failure hospitalization, and change from baseline in quality of life, sample size reassessment, p-value adjustment based on Cui-Hung-Wang method, O'Brien-Fleming boundary

Review file: c:\data\my ocuments\review\nda\bidil_A-HeFT.doc

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Based on A-HeFT, Bidil gives a statistically significantly better mean composite score than placebo in African-American patients with heart failure. The benefit of Bidil with respect to the composite score can be further explained by a statistically significantly smaller all-cause mortality rate ($p = 0.012$) and a statistically significantly smaller incidence rate of heart failure hospitalization ($p < 0.001$). The results of A-HeFT seem to provide support for the post hoc findings for black patients in V-HeFT I and II. In A-HeFT, the quality of life results show a trend in favor of Bidil but statistical significance is inconclusive (the pre-specified primary analysis gives $p = 0.24$).

A worrisome observation is that the Bidil's effect on mortality appears to be entirely contributed by the patients that are not analyzed in interim analysis #2 where a data-driven sample size increase takes place. However, the exploratory analysis results (Table 9 and Figure 3) do not provide sufficient explanation for this observation. According to the sponsor's analyses, possible differences in baseline covariates between the interim-analysis cohort and the post-interim-analysis cohort do not materially impact the estimated effect of Bidil and its statistical significance. Throughout the trial, usage of concomitant cardiovascular medications seems balanced between the treatment groups, except possibly lower usage of non-aldosterone antagonist diuretics in the patients receiving Bidil in the Look-2 cohort. Compared to the Look-2 cohort, the post-Look-2 cohort appears to have higher usage of ACE inhibitors or ARB and beta-blockers at Month 3. However, based on the sponsor's analysis adjusting for the six concomitant cardiovascular medication classes, the effect of Bidil does not appear to be significantly impacted.

1.2 Brief Overview of Clinical Studies

Bidil is a fixed-dose combination tablet of isosorbide dinitrate and hydralazine hydrochloride (ISDN+HYD). This combination was studied in the V-HeFT I and II trials in patients with heart failure. The V-HeFT results were submitted to the Agency in an NDA. The Agency and more than two Cardio-Renal Advisory Committees concluded that the V-HeFT results failed to provide conclusive evidence for the mortality benefit of this fixed-dose combination.

The V-HeFT I results seem to suggest a possible mortality benefit for ISDN+HYD compared to placebo ($p = 0.093$, logrank test; 27% reduction in mortality risk). V-HeFT II seems to suggest that this fixed-dose combination may have a smaller mortality benefit than enalapril ($p = 0.083$, logrank test; 23% larger mortality risk). According to the study report, analyses of data from the V-HeFT studies and from beta-blocker trials in heart failure suggest that when used for the treatment of HF, ACE inhibitors and beta-blockers may produce less clinical benefit in black patients than in non-black patients, whereas in contrast, ISDN+HYD appears to be more beneficial in black patients than in non-black patients, as suggested by the subgroup results below.

V-HeFT I results on all-cause mortality

[Source: excerpted from sponsor's Table 2, page 25, Volume 121.4]

	Blacks (N=180)			Whites (N=450)		
	ISDN+ HYD	Placebo	nominal p-value	ISDN+ HYD	Placebo	nominal p-value
Annual mortality rate (%)	9.7	17.3	0.04	16.9	18.8	ns
Mortality risk ratio	0.34	----	0.004	0.75	----	0.11

V-HeFT II results on all-cause mortality

[Source: excerpted from sponsor's Table 3, page 27, Volume 121.4]

	Blacks (N=215)			Whites (N=5740)		
	ISDN+ HYD	Enalapril	nominal p-value	ISDN+ HYD	Enalapril	nominal p-value
Annual mortality rate (%)	12.9	12.8	ns	14.9	11.0	0.02
Mortality risk ratio	0.95	----	0.83	1.48	----	0.009

Based on these results, A-HeFT which was conducted to demonstrate the safety and efficacy of Bidil in comparison to placebo in African-American patients with moderate to severe symptomatic heart failure while on standard therapy was intended to provide confirmatory data in support of the NDA. The primary efficacy endpoint was a composite score of clinical outcomes, calculated as the sum of the following components: death (score=-3), first heart failure hospitalization (score = -1), and change from baseline in quality of life at six months (score can be -2, -1, 0, 1, 2 from degree of worsening to degree of improvement). For example, if the patient had the quality of life score worsening greatly at six month, had been hospitalized for heart failure and died, then his/her composite score would be (-2)+(-1)+(-3) = -6. The secondary endpoints include the three component endpoints and five others. A-HeFT was terminated at the sample size of 1014, close to the proposed final sample size of 1100, mainly because of a significant mortality effect with Bidil. The final analysis was based on the total of 1050 patients that were randomized.

1.3 Statistical Issues and Findings

A-HeFT increased the initially planned sample size from 600 to 1100, partly based on the observed treatment difference at interim analysis #2 and according to a pre-specified algorithm agreed upon by the Agency. The p-value, adjusted for such an increase, of the composite score is $p = 0.021$ which is far smaller than the alpha level of 0.044, adjusted for the interim analyses and the sample size increase using the pre-specified adjustment approach of Cui, Hung and Wang (1999, Biometrics). Therefore, it can be concluded that Bidil yields a statistically significantly better mean composite score than placebo. The benefit of Bidil with respect to the composite score can be further explained by a statistically significantly smaller all-cause mortality rate ($p = 0.012$) and a statistically significantly smaller incidence rate of first failure hospitalization ($p<0.001$). The results of A-HeFT seem to provide support for the post hoc V-HeFT subgroup analysis findings in black patients.

The worrisome observation (Table 8) made by this reviewer is that the Bidil's effect on mortality appears to be entirely contributed by the patients that are not analyzed in interim analysis #2 where the sample size increase takes place. However, the results of Table 9 and Figure 3 do not seem to confirm this observation; at least, the apparently much larger reduction in mortality risk observed in the beginning of the post-interim-analysis-2 cohort does not persist throughout that cohort. According to the sponsor's analyses, possible differences in baseline covariates between the interim-analysis cohort and the post-interim-analysis do not materially impact the estimated effect of Bidil and its statistical significance. Throughout the trial, usage of concomitant cardiovascular medications seems balanced between the treatment groups, except possibly lower usage of non-aldosterone antagonist diuretics in the patients receiving Bidil in the Look-2 cohort. Compared to the Look-2 cohort, the post-Look-2 cohort appears to have higher usage of ACE inhibitors or ARB and beta-blockers at Month 3. However, based on the sponsor's analysis adjusting for the six concomitant cardiovascular medication classes, the effect of Bidil is not significantly impacted at Months 3, 6, 9. Data available for Months 12, 15 and 18 are too limited to provide accurate analyses.

The quality of life results show a trend in favor of Bidil, but statistical significance is inconclusive (the pre-specified primary analysis gives $p = 0.24$).

There was no statistically significant finding on any of the other five secondary endpoints.

2. INTRODUCTION

2.1 Overview

Bidil is a fixed-dose combination tablet of isosorbide dinitrate and hydralazine hydrochloride (ISDN+HYD). This combination was studied in the V-HeFT I and II trials in patients with heart failure (HF). The V-HeFT results were submitted to the Agency in an NDA. The Agency and more than two Cardio-Renal Advisory Committees concluded that the V-HeFT results failed to provide conclusive evidence for the mortality benefit of this fixed-dose combination.

The V-HeFT I results seem to suggest a possible mortality benefit for ISDN+HYD compared to placebo ($p = 0.093$, logrank test; 27% reduction in mortality risk). V-HeFT II seems to suggest that this fixed-dose combination may have a smaller mortality benefit than enalapril ($p = 0.083$, logrank test; 23% larger mortality risk). According to the study report, analyses of data from the V-HeFT studies and from beta-blocker trials in HF suggest that when used for the treatment of HF, ACE inhibitors and beta-blockers may produce less clinical benefit in black patients than in non-black patients, whereas in contrast, ISDN+HYD appears to be more beneficial in black patients than in non-black patients, as suggested by the subgroup results given below.

V-HeFT I results on all-cause mortality

[Source: excerpted from sponsor's Table 2, page 25, Volume 121.4]

	Blacks (N=180)			Whites (N=450)		
	ISDN+ HYD	Placebo	nominal p-value	ISDN+ HYD	Placebo	nominal p-value
Annual mortality rate (%)	9.7	17.3	0.04	16.9	18.8	ns
Mortality risk ratio	0.34	----	0.004	0.75	----	0.11

V-HeFT II results on all-cause mortality

[Source: excerpted from sponsor's Table 3, page 27, Volume 121.4]

	Blacks (N=215)			Whites (N=5740)		
	ISDN+ HYD	Enalapril	nominal p-value	ISDN+ HYD	Enalapril	nominal p-value
Annual mortality rate (%)	12.9	12.8	ns	14.9	11.0	0.02
Mortality risk ratio	0.95	----	0.83	1.48	----	0.009

Based on these results, A-HeFT which was conducted to demonstrate the safety and efficacy of Bidil in comparison to placebo in African-American patients with moderate to severe symptomatic heart failure while on standard therapy was intended to provide confirmatory data in support of the NDA. This statistical review pertains to this study.

2.2 Data Sources

SAS datasets in \\CDSESUB1\N20727\N_000\2004-10-29; a CD submitted on 12/21/2004.

3.1 Evaluation of Efficacy

A-HeFT was a randomized, double-blind, parallel group, placebo-controlled study of Bidil in stable, moderately to severely symptomatic (on standard heart failure therapy) African-American heart failure patients. The primary efficacy endpoint of A-HeFT was a composite score calculated from death, first hospitalization for heart failure, and change from baseline in six-month quality of life (QOL) measurement (assessed by Minnesota Living with Heart Failure questionnaire). According to the study report, first patient/first visit was on May 29, 2001 and the first patient was randomized on June 12, 2001. A Data and Safety Monitoring Committee (DSMB) monitored the safety of patients periodically during the trial. Subsequent to their meeting held on July 7 and 9, 2004, the DSMB recommended that A-HeFT be discontinued due to a statistically significant favorable mortality benefit of Bidil relative to placebo, based on analysis of survival in 1014 enrolled patients. After consultation with FDA, the sponsor informed the investigators of the study termination and requested that enrollment be stopped immediately on July 19, 2004. Last patient/last visit of the 1050 randomized patients was September 30, 2004. All patients were to be treated and followed for either a maximum of 18 months or until the last patient randomized had completed six months of treatment whichever occurred first. Because of the decision to terminate the study early, patients randomized in the later months of the trial did not complete the original intended minimum duration of 6 months.

DSMB was charged with the primary responsibilities to periodically review study results, evaluate the treatments for excess adverse effects, determine whether the basic trial assumptions remained valid, judge whether the overall integrity and conduct of the trial remained acceptable, and make recommendations to the sponsor and the A-HeFT Steering Committee. According to the study report, the DSMB reviewed the results of two interim analyses of the primary efficacy endpoint for the reassessment of sample size.

A total of 1,631 patients were entered into screening at 180 sites in the United States; 581 patients (36%) of screened patients were not randomized. Randomization was stratified for beta-blocker usage. The most frequent reasons for failure to complete the screening period were: failure to demonstrate symptomatic stability at screening (36%), failure to meet criteria for abnormal left ventricular function with 6 months of entry into screening (33%), change in body weight greater than 2.5% between screening and baseline (18%), change in signs and symptoms of heart failure or change in NYHA class between screening and baseline (11%). Patient disposition of the intent-to-treat, primary analysis population is summarized in Table 1.

Table 1. Disposition of patients (primary analysis population)
[Source: excerpted from Sponsor's Table 10]

	Bidil (N=518)	Placebo (N=532)
Number of patients randomized	518	532
Completers	469 (91%)	457 (86%)
discontinued study drug prematurely	153 (30%)	101 (19%)
Discontinued from study prematurely	49 (9%)	75 (14%)
Investigator decision	9 (2%)	13 (2%)
Patient withdrew consent	5 (1%)	3 (1%)
Lost to follow-up	2 (0%)	0 (0%)
Cardiac transplantation	3 (1%)	3 (1%)
Death	30 (6%)	54 (10%)
Not reported	0	2 (0%)
Final status for assessment of the composite endpoint		
Vital status known at study completion	518 (100%)	532 (100%)
Hospitalization status known at study completion	505 (98%)	521 (98%)
QOL assessment done at or before six-month visit	472 (91%)	497 (93%)

The percent of patients who took prohibited concomitant medications (hydralazine, long-acting nitrate, phosphodiesterase-5 inhibitors) appears to be similar in both treatment groups (Table 2).

Table 2. Number of percent of patients who took prohibited concomitant medication by treatment (safety population)
[Source: excerpted from Sponsor's Table 10]

	Bidil (N=518)	Placebo (N=532)
Number of patients who took prohibited concomitant medication	71 (14%)	90 (17%)
Hydralazine	14 (3%)	15 (3%)
Long-acting nitrate	65 (13%)	78 (15%)
Phosphodiesterase-5 inhibitor	3 (1%)	4 (1%)

The treatment groups appeared well-balanced on baseline characteristics, cardiovascular history and background therapies, except possibly on gender, mean blood pressure, diabetes mellitus, hyperlipidemia, and use of insulin and oral hypoglycermic drugs; see the column of 'entire population' in Table 3. There were more male patients in the placebo group (64%) than in the Bidil group (56%). The mean baseline blood pressures were 127/78 and 125/76 mm Hg in the Bidil and placebo groups, respectively.

The mean time on trial was longer for patients treated with Bidil (379 days) than those treated with placebo (355 days). The excess withdrawal from the trial in the placebo group was due to more patients withdrawn because of investigator decision (2.4% vs. 1.7%) and more deaths on the placebo group (10% vs., 6%). The duration of exposure to study drug was less in Bidil patients than in placebo patients (mean duration of 298 days vs. 314 days). This difference in exposure reflects the higher withdrawal rate on Bidil, which was primarily due to adverse events. Compliance appeared similar in both treatment groups (Sponsor's Table 21, page 99, Volume 121.4).

The pattern of concomitant cardiovascular medication use was similar to that observed for baseline medications, with the exception that the frequency of Natrecor use was much higher during the trial than at baseline. Percent of patients who used concomitant medications appeared similar in the two treatment groups (Table 3), except possibly on more frequent use of oral hypoglycemics in the Bidil group, more frequent use of antacids and propulsives in the placebo group (Sponsor's Table 17, page 94, Volume 121.4).

Table 3. Baseline characteristics, cardiovascular history, and background therapies
[Source: reviewer's Analysis]

	Look-2 cohort		post Look-2 cohort		Entire population	
	Bidil (N=164)	Placebo (N=152)	Bidil (N=354)	Placebo (N=380)	Bidil (N=518)	Placebo (N=532)
Gender						
Male	59.2%	66.5%	54.5%	62.9%	56.0%	63.9%
Female	40.8%	33.5%	45.5%	37.1%	44.0%	36.1%
Age (mean±sd)	56±12	56±14	57±13	57±13	57±13	57±13
< 65 yrs	73.2%	74.3%	68.4%	70.3%	69.9%	71.4%
≥ 65 yrs	26.8%	25.7%	31.6%	29.7%	30.1%	28.6%
Weight (kg)	91±27	94±25	92±25	94±26	92±26	94±25
Blood pressure						
Systolic	126±20	121±26	128±18	125±22	128±19	124±24
Diastolic	76±19	71±24	77±11	75±14	77±14	74±17
Heart rate	75±12	72±18	74±11	75±11	74±11	74±14
EF (%)	23.6±7.2	23.8±7.3	24.1±7.4	24.3±7.6	23.9±7.3	24.2±7.5
Hypertension	86.0%	86.8%	93.5%	88.4%	91.1%	88.0%
Arrhythmias	33.5%	35.5%	32.2%	34.2%	32.6%	34.6%
Diabetes Mellitus	40.2%	36.2%	46.9%	37.4%	44.8%	37.0%
Hyperlipidemia	45.7%	41.5%	60.5%	52.6%	55.8%	49.4%
Cerebrovascular disease	17.7%	17.1%	14.1%	12.6%	15.3%	13.9%
Peripheral vascular disease	12.8%	13.2%	10.5%	13.4%	11.2%	13.4%
COPD	20.1%	25.7%	16.4%	18.7%	17.6%	20.7%

Chronic renal insufficiency	15.9%	18.4%	16.4%	18.2%	16.2%	18.2%
Valvular disease	29.3%	30.3%	39.0%	39.0%	35.9%	36.5%
Previous implanatable pacemaker or ICD	14.6%	14.5%	17.5%	18.4%	16.6%	17.3%
Previous MI	28.7%	25.7%	29.7%	29.7%	29.3%	28.6%
Angina	0.6%	0.0%	0.6%	0.3%	0.6%	0.2%
Unstable angina in the past 3 months	0.0%	0.0%	0.3%	0.0%	0.2%	0.0%
Cigarette smoking during the past year	31.7%	25.7%	25.7%	26.6%	27.6%	26.3%
Previous cigarette smoking	62.8%	66.5%	57.1%	61.8%	58.9%	63.2%
Stroke	11.0%	11.2%	11.3%	10.0%	11.2%	10.3%
Atrial Fibrillation	18.9%	19.7%	13.8%	16.8%	15.4%	17.7%
TIA	6.7%	6.6%	3.4%	3.4%	4.4%	4.3%
Etiology of HF						
Ischemic	22.6%	22.4%	23.7%	22.9%	23.4%	22.7%
Idiopathic	25.0%	29.0%	24.3%	27.1%	24.5%	27.6%
Hypertensive	39.0%	36.2%	40.4%	37.9%	40.0%	37.4%
Valvular	3.7%	4.0%	2.0%	2.9%	2.5%	3.2%
others	9.8%	8.6%	9.6%	9.2%	9.7%	9.0%
Dyspnea						
Mild	25.6%	30.3%	26.8%	30.0%	26.5%	30.1%
Moderate	64.0%	57.2%	62.2%	55.5%	62.7%	56.0%
Severe	7.3%	7.9%	5.4%	8.4%	6.0%	8.3%
None	3.1%	4.6%	5.7%	6.1%	4.8%	5.6%
Orthopnea						
Mild	24.4%	32.9%	32.8%	34.5%	30.1%	34.0%
Moderate	37.2%	38.2%	38.1%	35.8%	37.8%	36.5%
Severe	11.6%	9.2%	7.3%	6.1%	8.7%	7.0%
None	26.8%	19.7%	21.5%	23.7%	23.2%	22.6%
Fatigue						
Mild	26.2%	23.0%	27.4%	29.8%	27.0%	27.8%
Moderate	61.6%	61.2%	57.6%	53.4%	58.9%	55.6%
Severe	8.5%	12.5%	11.0%	11.8%	10.2%	12.0%
None	3.1%	3.3%	4.0%	5.0%	3.7%	4.5%
Hospitalized in the past year for HF	92.7%	96.7%	61.3%	67.6%	71.2%	75.9%
NYHA class						
I	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
II	0.6%	0.0%	0.0%	0.0%	0.2%	0.0%
III	95.7%	92.8%	97.2%	95.5%	96.7%	94.7%
IV	3.7%	7.2%	2.8%	4.5%	3.1%	5.3%
ACE inhibitors	79.9%	77.0%	72.0%	74.5%	74.5%	75.2%
ARB	14.6%	16.5%	28.3%	22.9%	23.9%	21.1%
Beta blockers	76.2%	76.3%	87.3%	84.5%	83.8%	82.1%
Calcium blockers	18.3%	17.1%	22.3%	20.5%	21.0%	19.6%
Non-aldosterone antagonist diuretics	91.5%	95.4%	91.2%	91.8%	91.3%	92.9%
Aldosterone antagonist diuretics	40.2%	33.6%	40.1%	39.5%	40.2%	37.8%
Digitalis glycosides	70.1%	73.7%	53.4%	55.8%	58.7%	60.9%

Look-2 (interim analysis #2): sample size re-estimation occurred

There were ten amendments to the protocol. The sample size increase was in Protocol Amendment #8.

Primary efficacy endpoint

The primary efficacy endpoint was a composite score of clinical outcomes, calculated as the sum of the following components:

- death (adjudicated; all cause mortality)
  death at any time during the trial = -3
  alive at end of trial = 0
- first hospitalization for heart failure
  first hospitalization for heart failure (adjudicated) at any time during the trial = -1
  no hospitalization = 0
- change in overall score of QOL (Minnesota Living with Heart Failure questionnaire) at 6 months (or last available measurement if earlier than 6 months)
  improvement ≥ 10 units = +2
  improvement ≥ 5 and < 10 units = +1
  change less than 5 units (improvement or worsening) = 0
  worsening ≥ 5 and < 10 units = -1
  worsening ≥ 10 units = -2

Death

All-cause mortality was used in the primary efficacy analysis. Deaths occurring from any cause at any time in the trial were counted.

Hospitalization for heart failure

Occurrence of the first hospitalization for heart failure (or lack of any hospitalization for heart failure) at any time in the trial was counted.

Quality of Life (QOL)

The overall score for the Minnesota Living with Heart Failure questionnaire administered at six months (or the last available assessment before this if a six-month one was not available) was compared to baseline QOL to measure change in QOL for the composite primary endpoint.

Handling of Missing data

As stated in the protocol, missing data for any component was to be assigned the worst score in the primary analysis. This conservative approach was based on the anticipation that, in a trial of this duration, there would be little missing data and few patients lost to follow-up. For example, living patients who withdrew consent but agreed to allow their vital status to be recorded at

study end were assumed to have been hospitalized for heart failure (score = -1), but were recorded as alive (score = 0).

Interim analysis, sample size re-estimation and statistical analysis method

The sponsor articulated at the time of planning the sample size that there were no published data available that used the exact composite score endpoint proposed for this trial; therefore, the amount of variability for this composite score endpoint could not be determined precisely. Using previous data from studies such as V-HeFT II, this study was designed to be able to detect a between-group difference equivalent to 22.8% of a standard deviation with 300 patients per arm, at a two-sided $\alpha = 0.05$ and power 80%.

Two interim analyses were planned; the first (Look 1) when either 25% of the 600 patients had completed six months of follow-up or when 600 patients had been recruited, whichever occurred first and the second (Look 2) when 50% of the 600 patients had completed six months of follow-up. Since the objective of the interim analysis was to re-estimate the sample size and not to stop the trial early for significant efficacy, O'Brien-Fleming type boundaries were used. The two-sided p-values required for statistical significance were 0.00001 at Look 1, 0.0052 at Look 2, and 0.0480 at the final analysis.

The results of Look 2 were used to formally modify the study sample size. The sample size was re-estimated to provide 80% power to detect at a two-sided significance level 0.02 an effect size estimated by the observed mean difference divided by its standard deviation for the composite score endpoint using ANOVA at the time of the interim analysis. It was planned that the revised sample size could not be smaller than the originally planned 300 patients per treatment group and could not exceed 1000 patients per group. At the second interim analysis, it was decided to increase sample size to 550 per group (total sample size =1100). As the result of the re-estimation using the information from the interim sample path, the commonly used two-sample t statistic is no longer statistically valid. Many adaptive tests can be used. For the primary efficacy analysis, the sponsor chose to use the method of Cui, Hung and Wang (1999) for adjusting the two-sample t statistic with the statistical significance boundaries unchanged. All of these were pre-specified in the study protocol. The details on the sample size re-estimation criteria and the adaptive test are provided in Appendix 1.

After the DSMB's meetings were held on July 7 and 9, 2004, the DSMB unanimously recommended that A-HeFT be discontinued due to a statistically significant favorable mortality benefit of Bidil relative to placebo. At that time, 75 deaths were known to have occurred among 1014 patients, the number of patients randomized at the time the data was provided to the analysis statistician. The DSMB informed the sponsor that part of deliberations include use of a two-sided stopping rule for the log-rank test with O'Brien-Fleming boundaries at the 0.05 level for survival time. The DSMB also remarked that the mortality results were confirmed by and consistent with the composite score primary outcome, and its individual components, using data available at that time. The sponsor stopped A-HeFT on July 19, 2004. Investigators were instructed to stop enrollment and bring patients back to the site for a final visit. At this point, 1050 patients had been enrolled in A-HeFT, and of these, 952 patients had reached a minimum of 3 months on study and would have had the opportunity to participate in the first scheduled post baseline QOL measurement. 98 patients had not had the opportunity to complete any scheduled post-baseline QOL assessment; i.e., they had not been in the study for 3 months.

Results on Primary Efficacy Endpoint

According to the study report, at the conclusion of the trial, no patients were lost to follow-up for the assessment of vital status, 24 (2.3%) were lost to follow-up for the assessment of HF hospitalization and were assigned the worst score (-1), and 81 (7.7%) of patients had no QOL measurement done at or before their six-month visit and were assigned the worse score (-2) in the primary analysis.

Table 4 summarizes the results on the primary efficacy endpoint – composite score of death, first hospitalization for heart failure, and change from baseline in six-month quality of life. As mentioned above, because the sample size increase resulted from the observed treatment difference in this endpoint at Look 2, the unadjusted p-value of 0.011 can be misleading. The correct p-value is 0.021 according to the reviewer's analysis based on the protocol pre-specified CHW adjustment; the sponsor's adjusted p-value of 0.016 is incorrect because 32 patients contribute two data values in their calculation. Because the trial is stopped at the sample size 1014, not at the proposed final sample size of 1100, this analysis is an interim analysis although it is almost at the study end. The nominal alpha level for testing the primary efficacy endpoint at this analysis should be 0.044 (for more details, see Appendix 2), not 0.048 used by the sponsor. Nonetheless, since the adjusted p-value 0.021 is below 0.044, it can be concluded that Bidil yields a statistically significantly better mean composite score than placebo.

Table 4. Mean composite score of death, first hospitalization for heart failure, and change from baseline in six-month quality of life – primary analysis
[*Source: sponsor's analysis and reviewer's analysis]*

	Bidil (N=518)	Placebo (N=532)	p-value
Composite score	-0.16	-0.47	0.011[1] 0.016[2] 0.021[3]

[1] unadjusted two-sample t test
[2] sponsor's incorrect calculation using adaptive two-sample t test of Cui, Hung and Wang
[3] reviewer's calculation using adaptive two-sample t test of Cui, Hung and Wang

Four or five sensitivity analyses including LOCF and per protocol analyses were performed in the study protocol to assess the impact of handling missing values. The nominal p-values are from 0.001 to 0.014, except per-protocol analysis that gives p = 0.46.

Table 5 suggests that the beneficial effect of bidil on the composite score appears to be smaller in the patients analyzed in Look 2 where the sample size re-estimation occurred.

Table 5. Mean composite score of death, first hospitalization for heart failure, and change from baseline in six-month quality of life , by time of sample size re-estimation
[*Source: reviewer's analysis]*

	Look-2 cohort			post Look-2 cohort		
	Bidil (N=164)	Placebo (N=152)	Difference (B – P)	Bidil (N=354)	Placebo (N=380)	Difference (B – P)
Composite score	-0.23	-0.47	0.24	-0.07	-0.38	0.31

Look-2 (interim analysis #2): sample size re-estimation occurred

Secondary efficacy endpoints

A total of eight secondary endpoints are listed in the protocol. Of them, the three individual components (death, first hospitalization for heart failure, change from baseline in QOL at six months) of the primary efficacy endpoint automatically have to be analyzed, regardless of whether other secondary endpoints are listed or not; thus, these component endpoints should be treated differently from others in assessment of statistical significance.

The results of the three individual components based on the mean score analysis are summarized in Table 6. The Bidil group appears to show a statistically significantly better mean score on death and first HF hospitalization than the placebo group, but not on the QOL component.

Table 6. Mean scores of individual components of the composite score of death, first hospitalization for heart failure, and change from baseline in six-month quality of life – primary analysis
[Source: sponsor's analysis and reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	p-value[1]
Death	-0.19	-0.30	0.019
First hospitalization for heart failure	-0.19	-0.27	0.003
Change from baseline in QOL at 6 months	0.21	0.10	0.24

[1] two-sample t test

Death and First Heart Failure Hospitalization

Death and first HF hospitalization were also analyzed by time to event methods as secondary efficacy analysis, as shown in Table 7. The Bidil group seemed to have a longer time on trial (Appendix 3), which may result in underestimation of the effect of Bidil. Based on the consistent results in Tables 6 and 7, it can be concluded that Bidil yields a statistically significant benefit on

all-cause mortality and on heart failure hospitalization. Based on Figures 1 and 2, the effect of Bidil on mortality or HF hospitalization does not appear to be constant over time (also see Figures A4-1 and A4-2 in Appendix 4).

Table 7. Event rate and time to event analysis for deaths and hospitalization for heart failure - secondary efficacy analysis
[Source: sponsor's analysis and reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)	p-value[1]
Death	32 (6.2%)	54 (10.2%)	0.57 (0.37, 0.89)	0.012
First hospitalization for heart failure	85 (16.4%)	130 (24.4%)	0.61 (0.46, 0.80)	< 0.001

[1] Cox regression analysis

Figure 1. Kaplan-Meier estimates for all-cause mortality by treatment

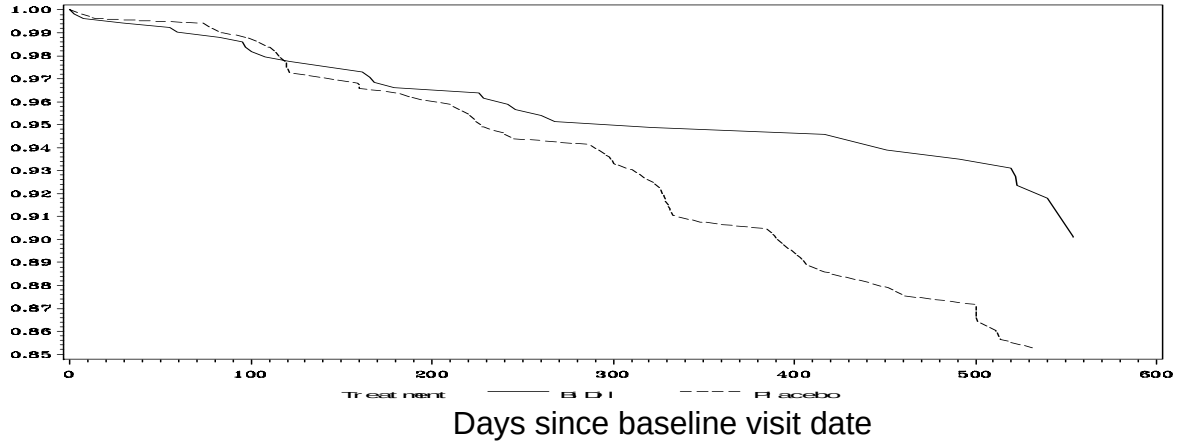
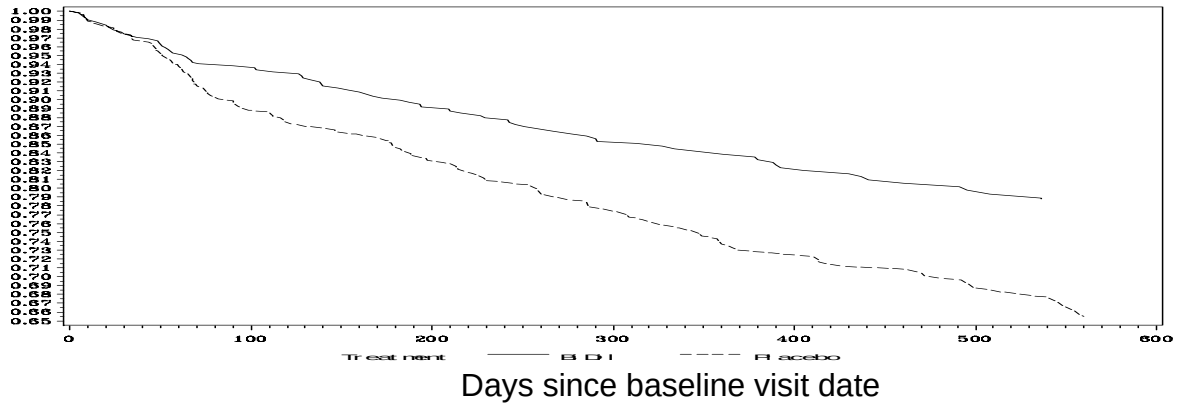


Figure 2. Kaplan-Meier estimates for HF hospitalization by treatment



Due to the sample size increase based on the observed treatment difference in Interim Analysis #2, this reviewer examined whether there are substantial differences in the efficacy results between the Look-2 cohort that include patients that were analyzed in the interim analysis #2 and the post Look-2 cohort of the remaining patients. The results are summarized in Table 8. For all-cause mortality, the Bidil's benefit seems to be very much smaller in the Look-2 cohort (7% reduction of mortality risk) as compared to the post Look-2 cohort (62% reduction). For first HF hospitalization, the bidil's beneficial effect is also smaller in the Look-2 cohort, but to a much lesser extent than for the mortality.

Table 8. Event rate and time to event analysis for deaths and hospitalization for heart failure – by time of sample size re-estimation
[Source: reviewer's analysis]

	Look-2 cohort			post Look-2 cohort		
	Bidil (N=164)	Placebo (N=152)	HR (95% CI)	Bidil (N=354)	Placebo (N=380)	HR (95% CI)
Death	18 (11.0%)	18 (11.8%)	0.93 (0.49, 1.79)	14 (4.0%)	36 (9.5%)	0.38 (0.21, 0.71)
First HF hospitalization	35 (21.3%)	48 (31.6%)	0.66 (0.42, 1.01)	50 (14.1%)	82 (21.6%)	0.58 (0.41, 0.82)

Look 2 (interim analysis #2): sample size re-estimation occurred

To explore possible reasons for the inconsistency, a number of additional analyses were performed. Table 3 gives a comparison between the Look-2 cohort and the post Look-2 cohort with respect to baseline characteristics, cardiovascular history, and background therapies. The percent of patients who had been hospitalized in the past year for heart failure before randomization was quite different between the two cohorts. The sponsor explained in the 3/23/2005 document that this difference occurred because the study protocol was amended on March 25, 2003 to eliminate the requirement for a prior hospitalization for heart failure in order to enhance enrollment (since the sample size needs to increase by nearly 100%). And because it was based on changing health care delivery patterns, the elimination of the prior hospitalization criterion was not expected to change the clinical status of the A-HeFT patients. Furthermore, the sponsor cited a number of references to suggest that the influence of previous hospitalizations on subsequent hospitalization rate was little according to the COPERNICUS data.

A troublesome observation in Table 8 made by this reviewer is that the Bidil's effect on mortality appears to be entirely contributed by the patients that are not analyzed in interim analysis #2 where the sample size increase takes place. Table 9 presents the hazard ratios for every 100 patients enrolled subsequently and Figure 3 illustrates the result from the analysis of a moving window of 100 patients in steps of 10 patients. Interestingly, from Table 9, in the post-Look-2 cohort, the number of deaths in the placebo group doubled in the 4^{th} 100 patients and the 5^{th} 100 patients, whereas the number of deaths in the Bidil group is less than half. However, the apparently much larger reduction in mortality risk observed in the beginning of the post interim analysis #2 cohort does not persist throughout that cohort. The same analyses were performed on HF hospitalization (Table 10 and Figure 4).

Table 9. Number of deaths / sample size (%) for every 100 patients

[Source: reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio
1st 100 patients	7/48 (14.6%)	5/52 (9.6%)	1.49
2nd 100 patients	5/59 (8.5%)	7/41 (17.1%)	0.50
3rd 100 patients	5/50 (10.0%)	2/50 (4.0%)	2.65
4th 100 patients	1/49 (2.0%)	10/51 (19.6%)	0.09
5th 100 patients	3/46 (6.5%)	11/54 (20.4%)	0.31
6th 100 patients	3/50 (6.0%)	6/50 (12.0%)	0.50
7th 100 patients	2/52 (3.9%)	6/48 (12.5%)	0.29
8th 100 patients	2/50 (4.0%)	3/50 (6.0%)	0.63
9th 100 patients	2/47 (4.3%)	3/53 (5.7%)	0.71
remaining 150 patients	2/67 (3.0%)	1/83 (1.2%)	2.23

Figure 3. Moving window of 100 patients in steps of 10 patients for all-cause mortality
[Source: reviewer's analysis]

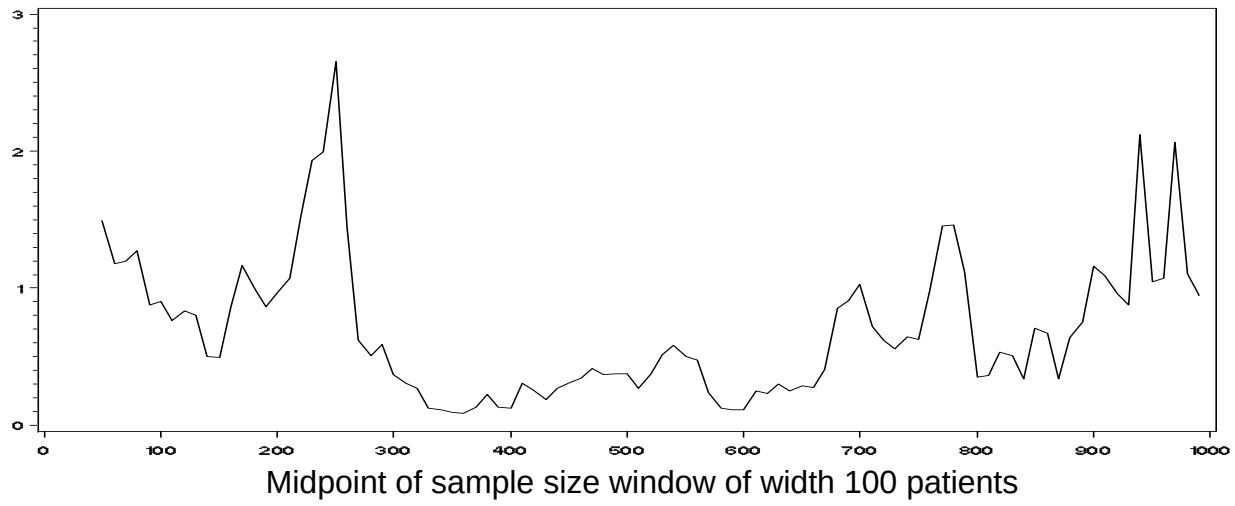
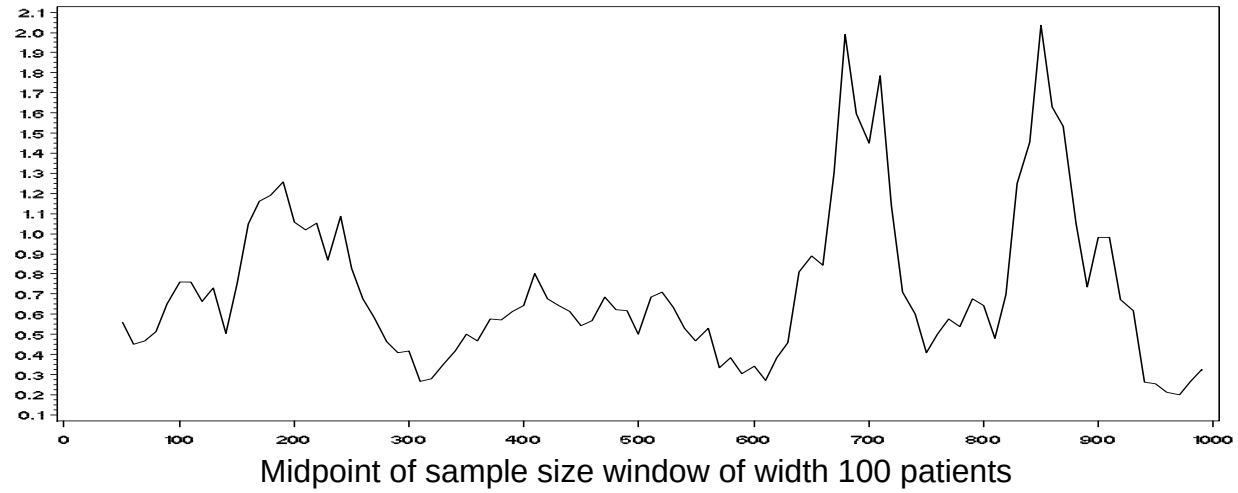


Table 10. Number of patients hospitalized for heart failure at least once / sample size (%) for every 100 patients *[Source: reviewer's analysis]*

	Bidil (N=518)	Placebo (N=532)	Hazard ratio
1st 100 patients	13/48 (27.1%)	23/52 (44.2%)	0.56
2nd 100 patients	11/59 (18.6%)	10/41 (24.4%)	0.75
3rd 100 patients	11/50 (22.0%)	14/50 (28.0%)	0.83
4th 100 patients	10/49 (20.4%)	17/51 (33.3%)	0.50
5th 100 patients	11/46 (23.9%)	22/54 (40.7%)	0.54
6th 100 patients	7/50 (14.0%)	14/50 (28.0%)	0.47
7th 100 patients	8/52 (15.4%)	8/48 (16.7%)	0.89
8th 100 patients	6/50 (12.0%)	13/50 (26.0%)	0.41
9th 100 patients	7/47 (14.9%)	4/53 (7.6%)	2.03
remaining 150 patients	1/67 (1.5%)	5/83 (6.0%)	0.21

Figure 4. Moving window of 100 patients in steps of 10 patients for all-cause mortality *[Source: reviewer's analysis]*



After being informed of this troublesome observation for mortality, the sponsor also performed a number of additional analyses to look into possible reasons for this apparent difference between the Look-2 cohort and the post-Look-2 cohort. First, inclusion of the indicator of the post-Look-2 cohort is not predictive for mortality (p-value for this indicator variable is 0.96). Treatment effect on hazard ratio and its p-value are virtually unchanged when this indicator variable is included in the Cox regression analysis. Secondly, analyses are performed to explore possible differences in baseline characteristics, cardiovascular history, baseline therapies, and time from diagnosis of heart failure to randomization, etc. Among all the baseline parameters assessed, only four parameters appear to be significantly different between the two cohorts; they are

baseline systolic blood pressure (p=0.0058), baseline diastolic blood pressure (p = 0.044), usage of beta blocker at baseline (p = 0.0001) and usage of digitalis glycoside at baseline (p < 0.0001). Any or all of these parameters are then included in conjunction with the post-Look-2 indicator variable in the Cox regression analyses. The results demonstrate that the effect of Bidil on the mortality in terms of hazard ratio and its p-value range from (HR = 0.611, p = 0.028) to (HR= 0.607, p = 0.026), which do not appear to be very different from those (HR = 0.571, p = 0.012) obtained from the original primary analysis.

Moreover, in response to this reviewer's request for examining possible differences in the use of concomitant cardiovascular medications at 3-month intervals during the course of the trial, the sponsor submitted their analysis results as summarized in Table 11. As usually suspected, a large number of patients are not accounted for in the analysis; the sample size decreases substantially from Month 3 to Month 18. Based on the sponsor's analysis, the post-Look-2 cohort appears to have higher usage of ACE inhibitors or ARB and beta-blocker at Month 3 and lower digitalis glycoside usage at Months 3, 6, 12, and 15. Between the treatment groups, there appears to be lower usage of non-aldosterone antagonist diuretics in the patients receiving Bidil in the Look-2 cohort; the difference is observed at Months 3, 6, 9, 12 and 15. The sponsor's further analysis by adjusting for usage of the six classes of concomitant medication via Cox regression model appears to suggest that the effect of Bidil is not significantly impacted by inclusion of these covariates at Months, 3, 6, 9 (Sponsor's Table FDA Qu Set 4 in the 04/26/2005 document). Data available for Months 12, 15 and 18 are too limited to provide accurate analyses.

Table 11. Use of concomitant medications during the post-randomization phase
[Source: excerpted from Sponsor's analysis]

	Look-2 cohort		post Look-2 cohort	
	Bidil	Placebo	Bidil	Placebo
Month 3				
	N=154	N=150	N=313	N=314
ACE-I or ARB	88.3%	90.7%	92.7%	94.3%
Beta blockers	76.6%	80.7%	86.9%	86.3%
Calcium blockers	18.8%	19.3%	20.8%	20.4%
Digitalis glycosides	70.1%	71.3%	54.0%	56.7%
Diuretics: Aldosternone	43.5%	36.0%	41.5%	41.7%
Diuretics: other	89.6%	96.7%	93.9%	91.7%
Month 6				
	N=149	N=142	N=255	N=250
ACE-I or ARB	88.6%	90.8%	92.2%	92.4%
Beta blockers	78.5%	83.8%	85.5%	85.6%
Calcium blockers	17.4%	21.1%	21.6%	23.6%
Digitalis glycosides	71.1%	74.6%	51.8%	54.0%
Diuretics: Aldosternone	41.6%	36.6%	42.4%	44.8%
Diuretics: other	89.9%	96.5%	93.7%	92.4%
Month 9				
	N=146	N=139	N=208	N=207
ACE-I or ARB	91.1%	87.8%	92.8%	91.8%
Beta blockers	77.4%	84.2%	86.1%	86.0%
Calcium blockers	19.2%	23.0%	21.2%	24,2%

Digitalis glycosides	68.5%	73.4%	51.4%	55.1%
Diuretics: Aldosternone	40.4%	37.4%	42.8%	41.1%
Diuretics: other	87.0%	96.4%	94.2%	92.3%
Month 12				
	N=145	N=130	N=160	N=155
ACE-I or ARB	91.7%	86.9%	91.9%	91.6%
Beta blockers	80.7%	86.2%	87.5%	82.6%
Calcium blockers	20.7%	20.8%	18.8%	25.8%
Digitalis glycosides	68.3%	73.8%	53.8%	56.8%
Diuretics: Aldosternone	43.4%	37.7%	38.1%	41.9%
Diuretics: other	86.9%	96.2%	90.6%	90.3%
Month 15				
	N=142	N=123	N=102	N=104
ACE-I or ARB	89.4%	87.8%	94.1%	88.5%
Beta blockers	79.6%	84.6%	89.2%	83.7%
Calcium blockers	22.5%	22.0%	23.5%	25.0%
Digitalis glycosides	66.2%	69.9%	52.9%	55.8%
Diuretics: Aldosternone	42.3%	35.8%	38.2%	39.4%
Diuretics: other	85.9%	95.9%	92.2%	92.3%
Month 18				
	N=57	N=44	N=37	N=34
ACE-I or ARB	93.0%	86.4%	97.3%	85.3%
Beta blockers	80.7%	90.9%	86.5%	79.4%
Calcium blockers	28.1%	22.7%	29.7%	23.5%
Digitalis glycosides	64.9%	56.8%	48.6%	50.0%
Diuretics: Aldosternone	45.6%	34.1%	45.9%	38.2%
Diuretics: other	91.2%	97.7%	94.6%	91.2%

Look-2 (interim analysis #2): sample size re-estimation occurred

Table 12 summarizes causes of deaths. Most of the deaths are heart failure deaths, according to the ICAC adjudication results.

The study report presents the results of all-cause hospitalization data that were analyzed in many ways – by event rate and by total days in hospital; see Table 13. Bidil does not seem to have a beneficial effect on the incidence rate of hospitalization or days in hospitalization.

Table 12. Number (%) of patients who died with causes of death (ICAC-adjudicated)
[Source: excerpted from sponsor's Table 31]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)
Death	32 (6.2%)	54 (10.2%)	0.57 (0.37, 0.89)
Heart failure deaths	21 (4.1%)	42 (7.9%)	0.61 (0.46, 0.80)
Sudden cardiac death	17 (3.3%)	24 (4.5%)	
Pump failure death	4 (0.8%)	16 (3.0%)	
MI-related death	0 (0.0%)	2 (0.4%)	
Cardiac procedure-related death	0 (0.0%)	0 (0.0%)	
Other cardiac cause-related death	0 (0.0%)	0 (0.0%)	
Non-heart failure (vascular death)	5 (1.0%)	3 (0.6%)	
Cerebrovascular accident death	4 (0.8%)	3 (0.6%)	
Vascular-related death	1 (0.2%)	0 (0.0%)	
Pulmonary embolism-related death	0 (0.0%)	0 (0.0%)	
Other vascular cause-related death	0 (0.0%)	0 (0.0%)	
Non-cardiovascular death	6 (1.2%)	9 (1.7%)	
Non-cardiovascular cause death	3 (0.6%)	5 (0.9%)	
Unknown cause death	3 (0.6%)	4 (0.8%)	

Table 13. All cause hospitalization event rates and total days in hospital
[Source: excerpted from sponsor's Tables 32 & 33]

	Bidil (N=518)	Placebo (N=532)	p-value
Event rate for hospitalization			
HF hospitalization	85 (16.4%)	130 (24.4%)	< 0.001#
All cause hospitalization	202 (39.0%)	221 (41.5%)	0.41$
Other cardiac cause hospitalization	80 (15.4%)	90 (16.9%)	0.56$
Non-cardiac cause hospitalization	109 (21.0%)	117 (22.0%)	0.76$
Days in hospital (days/patient)			
HF hospitalization			
Mean (SD)	13.7 (16.6)	15.3 (20.2)	0.54*
Range	2 - 122	2 - 164	
All cause hospitalization			
Mean (SD)	13.0 (15.6)	17.7 (21.6)	0.012*
Range	2 - 135	2 - 196	
Other cardiac cause hospitalization			
Mean (SD)	7.2 (10.0)	7.4 (5.7)	0.90*
Range	2 - 84	2 - 26	
Non-cardiac cause hospitalization			
Mean (SD)	8.1 (6.8)	10.6 (11.8)	0.051*
Range	2 – 34	2 - 65	

log-rank test $ Fisher's exact test * two-sample t test

Quality of Life

The pre-specified primary analysis for QOL as summarized in Table 6 does not demonstrate a statistically significant benefit on QOL with bidil. The study report includes the results of a number of additional analyses. Table 14 summarizes the result of the last available data analysis on Quality of Life assessed by the Minnesota Living with Heart Failure questionnaire.

Table 14. Mean change from baseline in overall, emotional and physical scores in Minnesota Living with Heart Failure questionnaire at endpoint
[Source: excerpted from sponsor's Table 36]

	Bidil (N=518)	Placebo (N=532)	p-value[1]
Overall score			
Mean baseline	50.9	50.8	
Mean change (SD)	-7.6 (22.6)	-3.4 (22.7)	0.003
Range of change	-91 – 68	-105 – 70	
Physical score			
Mean baseline	22.1	22.0	
Mean change (SD)	-3.5 (10.5)	-1.4 (10.6)	0.002
Range of change	-40 – 29	-401 – 30	
Emotional score			
Mean baseline	10.4	10.4	
Mean change (SD)	-1.3 (6.8)	-0.7 (6.5)	0.13
Range of change	-25 – 22	-25 – 17	

[1] two-sample t-test

Other Secondary endpoints

There was virtually no difference between the two treatment groups on the percent of patients assessed as needing cardiac transplantation during the study, the number of emergency room visits or unscheduled office/clinical visits for heart failure. Numerically, Bidil seems to be slightly favorable with respect to percent of patients with improvement or deterioration on NYHA functional class at six months or at endpoint but the differences are not statistically significant (Sponsor's Table 40, page 125, Volume 121.4).

Post Hoc Added Composite Endpoints

The study report also adds the results of two composite secondary endpoints; see Table 15. These analyses provide little additional information on death or hospitalization. There is no statistically significant effect of bidil on the composite endpoint of all-cause death or all-cause hospitalization.

Table 15. Event rate and time to event analysis for all-cause deaths and hospitalization – post hoc added secondary efficacy analysis
[Source: sponsor's analysis and reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)	p-value[1]
First hospitalization for heart failure or all-cause mortality	108 (20.8%)	158 (29.7%)	0.63 (0.49, 0.81)	< 0.001
All-cause hospitalization or all-cause mortality	215 (41.5%)	237 (44.5%)	0.86 (0.72, 1.04)	0.12

[1] Cox regression analysis

Reviewer's Additional Analyses

Randomization was stratified by beta-blocker usage at baseline. The reviewer's analysis using beta blocker usage as a stratification factor gives very similar results (p = 0.013 for composite score, stratified log rank p = 0.014 for all-cause death, stratified log rank p = 0.0004 for HF hospitalization) as what the unstratified analysis shows.

Dr. Lemtouni, Medical Reviewer, observed some possible imbalances in demographic characteristics, cardiovascular history, and background therapies between the two treatment groups. For her request, this reviewer performed a number of analyses adjusted for these covariates identified by her. The covariate analyses change little on the results.

3.2 Evaluation of Safety

Please read Dr. Lemtouni's review for safety assessment.

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

The bidil effects on the composite score, all cause mortality and HF hospitalization appeared to be consistent across gender or age subgroups.

Table 16. Subgroup analysis of primary efficacy endpoint – composite score

	Bidil (N=518)		Placebo (N=532)		Mean difference (95% CI)
	n	mean	n	mean	
Male	290	-0.28	340	-0.55	0.27 (-0.04, 0.57)
Female	228	-0.00	192	-0.33	0.33 (-0.06, 0.72)
< 65 yrs	362	-0.17	380	-0.34	0.17 (-0.11, 0.45)
≥ 65 yrs	156	-0.13	152	-0.79	0.65 (0.19, 1.12)

Table 17. Subgroup analysis of mortality

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)
Male	22/290 (7.6%)	32/340 (9.4%)	0.79 (0.46, 1.35)
Female	10/228 (4.4%)	22/192 (11.5%)	0.36 (0.16, 0.71)
< 65 yrs	24/362 (6.6%)	33/380 (8.7%)	0.73 (0.43, 1.24)
≥ 65 yrs	8/156 (5.1%)	21/152 (13.8%)	0.33 (0.15, 0.74)

Table 18. Subgroup analysis of HF hospitalization

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)
Male	46/290 (15.9%)	85/340 (25.0%)	0.60 (0.42, 0.86)
Female	39/228 (17.1%)	45/192 (23.4%)	0.62 (0.40, 0.95)
< 65 yrs	56/362 (15.5%)	93/380 (24.5%)	0.57 (0.41, 0.79)
≥ 65 yrs	29/156 (18.6%)	37/152 (24.3%)	0.70 (0.43, 1.41)

4.2 Other Special/Subgroup Populations

The bidil effects on all cause mortality and HF hospitalization appeared to be consistent across all subgroups, except possibly in small subgroups.

Table 19. Mortality for other subgroups
[Source: Sponsor's analysis and reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)
ACE-I			
Yes	25/386 (6.5%)	32/400 (8.0%)	0.78 (0.46, 1.31)
No	7/132 (5.3%)	22/132 (16.7%)	0.28 (0.12, 0.65)
ARB			
Yes	7/124 (5.6%)	14/112 (12.5%)	0.42 (0.17, 1.04)
No	25/394 (6.3%)	40/420 (9.5%)	0.62 (0.38, 1.03)
Beta blockers			
Yes	22/434 (5.1%)	32/437 (7.3%)	0.66 (0.38, 1.13)
No	10/84 (11.9%)	22/95 (23.2%)	0.46 (0.22, 0.97)
Calcium channel blockers			

Yes	8/109 (7.3%)	7/104 (6.7%)	1.07 (0.39, 2.94)
No	24/409 (5.9%)	47/428 (11.0%)	0.50 (0.30, 0.81)
Aldosterone Antagonist			
Yes	7/208 (3.4%)	22/201 (10.9%)	0.28 (0.12, 0.66)
No	25/310 (8.1%)	32/331 (9.7%)	0.80 (0.47, 1.35)
Non-aldosterone antag diuretic			
Yes	29/473 (6.1%)	52/494 (10.5%)	0.55 (0.35, 0.86)
No	3/45 (6.7%)	2/38 (5.3%)	1.13 (0.19, 6.79)
Digitalis glycoside			
Yes	20/304 (6.6%)	40/324 (12.3%)	0.50 (0.29, 0.86)
No	12/214 (5.6%)	14/208 (6.7%)	0.79 (0.36, 1.70)
History of hypertension			
Yes	25/472 (5.3%)	47/468 (10.0%)	0.49 (0.30, 0.80)
No	7/46 (15.2%)	7/64 (10.9%)	1.32 (0.46, 3.76)
Diabetes mellitus			
Yes	13/232 (5.6%)	18/197 (9.1%)	0.56 (0.28, 1.15)
No	19/286 (6.6%)	36/335 (10.7%)	0.59 (0.34, 1.03)
Chronic renal insufficiency			
Yes	6/84 (7.1%)	16/97 (16.5%)	0.44 (0.17, 1.11)
No	26/434 (6.0%)	38/435 (8.7%)	0.64 (0.39, 1.05)
Etiology of HF: Ischemic			
Yes	9/121 (7.4%)	18/121 (14.9%)	0.46 (0.21, 1.02)
No	23/397 (5.8%)	36/411 (8.8%)	0.63 (0.37, 1.06)
Etiology of HF: Hypertensive			
Yes	10/207 (4.8%)	18/199 (9.1%)	0.50 (0.23, 1.08)
No	22/311 (7.1%)	36/333 (10.8%)	0.61 (0.36, 1.04)
Baseline systolic BP			
> 125 mmHg	14/280 (5.0%)	15/247 (6.1%)	0.75 (0.36, 1.55)
≤ 125 mmHg	18/238 (7.6%)	39/285 (13.7%)	0.54 (0.31, 0.93)

Table 20. HF hospitalization for other subgroups
[Source: Reviewer's analysis]

	Bidil (N=518)	Placebo (N=532)	Hazard ratio (95% CI)
ACE-I			
Yes	65/386 (16.8%)	97/400 (24.3%)	0.65 (0.47, 0.89)
No	20/132 (15.2%)	33/132 (25.0%)	0.50 (0.29, 0.87)
ARB			
Yes	17/124 (13.7%)	24/112 (21.4%)	0.60 (0.32, 1.12)
No	68/394 (17.3%)	106/420(25.2%)	0.61 (0.45, 0.83)
Beta blockers			
Yes	66/434 (15.2%)	100/437(22.9%)	0.60 (0.44, 0.82)
No	19/84 (22.6%)	30/95 (31.6%)	0.64 (0.36, 1.14)
Calcium channel blockers			

Yes	17/109 (15.6%)	22/104 (21.2%)	0.71 (0.38, 1.34)
No	68/409 (16.6%)	108/428(25.2%)	0.59 (0.44, 0.80)
Aldosterone Antagonist			
Yes	30/208 (14.4%)	52/201 (25.9%)	0.48 (0.31, 0.76)
No	55/310 (17.7%)	78/331 (23.6%)	0.70 (0.50, 0.99)
Non-aldosterone antag diuretic			
Yes	80/473 (16.9%)	127/494(25.7%)	0.59 (0.45, 0.78)
No	5/45 (11.1%)	3/38 (7.9%)	1.39 (0.33, 5.83)
Digitalis glycoside			
Yes	60/304 (19.7%)	90/324 (27.8%)	0.65 (0.47, 0.90)
No	25/214 (11.7%)	40/208 (19.2%)	0.55 (0.33, 0.90)
History of hypertension			
Yes	77/472(16.3%)	107/468(22.9%)	0.65 (0.49, 0.88)
No	8/46 (17.4%)	23/64 (35.9%)	0.40 (0.18, 0.89)
Diabetes mellitus			
Yes	45/232 (19.0%)	50/197 (25.4%)	0.68 (0.46, 1.02)
No	40/286 (14.0%)	80/335 (23.9%)	0.53 (0.36, 0.78)
Chronic renal insufficiency			
Yes	22/84 (26.2%)	28/97 (28.9%)	0.89 (0.51, 1.55)
No	63/434 (14.5%)	102/435(23.5%)	0.55 (0.40, 0.76)
Etiology of HF: Ischemic			
Yes	21/121 (17.4%)	34/121 (28.1%)	0.54 (0.31, 0.93)
No	64/397 (16.1%)	96/411 (23.4%)	0.63 (0.46, 0.87)
Etiology of HF: Hypertensive			
Yes	24/207 (11.6%)	38/199 (19.1%)	0.56 (0.34, 0.94)
No	61/311 (19.6%)	92/333 (27.6%)	0.64 (0.46, 0.88)
Baseline systolic BP			
> 125 mmHg	42/280 (15.0%)	48/247 (19.4%)	0.70 (0.46, 1.06)
≤ 125 mmHg	43/238 (18.1%)	82/285 (28.8%)	0.57 (0.40, 0.83)

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

A-HeFT increased the initially planned sample size from 600 to 1100, partly based on the observed treatment difference at interim analysis #2 and according to a pre-specified algorithm agreed upon by the Agency. The p-value, adjusted for such an increase, of the composite score is $p = 0.021$ which is far smaller than the alpha level of 0.044, adjusted for the interim analyses and the sample size increase using the pre-specified adjustment approach of Cui, Hung and Wang (1999, Biometrics). Therefore, it can be concluded that Bidil yields a statistically significantly better mean composite score than placebo. The benefit of Bidil with respect to the composite score can be further explained by a statistically significantly smaller all-cause mortality rate ($p = 0.012$) and a statistically significantly smaller incidence rate of first failure hospitalization

($p<0.001$). The results of A-HeFT seem to provide support for the post hoc V-HeFT subgroup analysis findings in black patients.

The worrisome observation (Table 8) made by this reviewer is that the Bidil's effect on mortality appears to be entirely contributed by the patients that are not analyzed in interim analysis #2 where the sample size increase takes place. However, the results of Table 9 and Figure 3 do not seem to confirm this observation; at least, the apparently much larger reduction in mortality risk observed in the beginning of the post interim analysis #2 cohort does not persist throughout that cohort. According to the sponsor's analyses, possible differences in baseline covariates between the interim-analysis cohort and the post-interim-analysis cohort do not materially impact the estimated effect of Bidil and its statistical significance. Throughout the trial, usage of concomitant cardiovascular medications seems balanced between the treatment groups, except possibly lower usage of non-aldosterone antagonist diuretics in the patients receiving Bidil in the Look-2 cohort. Compared to the Look-2 cohort, the post-Look-2 cohort appears to have higher usage of ACE inhibitors or ARB and beta-blockers at Month 3. However, based on the sponsor's analysis adjusting for the six concomitant cardiovascular medication classes, the effect of Bidil is not significantly impacted at Months 3, 6, 9. Data available for Months 12, 15 and 18 are too limited to provide accurate analyses.

5.2 Conclusions and Recommendations

Bidil gives a statistically significantly better mean composite score than placebo. The benefit of Bidil with respect to the composite score can be further explained by a statistically significantly smaller all-cause mortality rate ($p = 0.012$) and a statistically significantly smaller incidence of HF hospitalization ($p<0.001$). The results of A-HeFT seem to provide support for the post hoc findings for black patients in V-HeFT I and II.

A worrisome observation is that the Bidil's effect on mortality appears to be entirely contributed by the patients that are not analyzed in interim analysis #2 where the data-dependent sample size increase takes place. However, the exploratory analyses (Table 9 and Figure 3) do not provide sufficient explanation for this observation. According to the sponsor's analyses, possible differences in baseline covariates between the interim-analysis cohort and the post-interim-analysis cohort do not materially impact the estimated effect of Bidil and its statistical significance. Throughout the trial, usage of concomitant cardiovascular medications seems balanced between the treatment groups, except possibly lower usage of non-aldosterone antagonist diuretics in the patients receiving Bidil in the Look-2 cohort. Compared to the Look-2 cohort, the post-Look-2 cohort appears to have higher usage of ACE inhibitors or ARB and beta-blockers at Month 3. However, based on the sponsor's analysis adjusting for the six concomitant cardiovascular medication classes, the effect of Bidil does not appear to be significantly impacted.

The quality of life results show a trend in favor of Bidil but statistical significance is inconclusive (the pre-specified primary analysis gives $p = 0.24$).

Appendix 1

The sponsor's proposed sample size increase algorithm and the adjustment procedure for the two-sample t test for the primary efficacy endpoint (composite score) are described as follows.

The initial planned sample is 600 patients in total (300 per treatment group). The second interim analysis (Look 2) was to occur when 50 of the 600 patients had completed six months of follow-up. The sample size was estimated at Look 2 to provide 80% power to detect the observed effect size at this look at two-sided significance level 0.02 but it would not exceed 1000 per treatment group. For testing the primary efficacy endpoint, because of this data-dependent sample size increase, the protocol specifies that the needed conservative adjustment on any p-value or test will be based on the procedure of Cui, Hung and Wang (1999, Biometrics).

Appendix 2

Since the sample size in the final analysis is 1050, not the proposed final sample size of 1100, the information time should be calculated as follows. Using the formula (3.1) of Cui, Hung and Wang (1999, Biometrics), we can calculate

$$b = (1100 – 300)/(600 – 300)$$

and the information time for 1050 on the original information time scale

$$t = \{(1050 – 300)/b + 300\}/600 = 0.96875.$$

Based on the FORTRAN program by Robussin et al, 2001 and available on http://www.medsch.wisc.edu/landemets/, the critical value at this information time is 2.0094 and the corresponding nominal alpha level is 0.044.

Appendix 3

Table A3-1. Censoring time distribution for survival time in patients who survived
[Source: Reviewer's Analysis]

	Look-2 cohort		post Look-2 cohort		Entire population	
In days	Bidil (N=146)	Placebo (N=134)	Bidil (N=340)	Placebo (N=344)	Bidil (N=486)	Placebo (N=478)
Mean	571	568	345	323	413	391
Standard deviation	104	81	178	188	190	199
Range	451-1111	393-1030	17-805	6-734	17-1111	6-1030
Percentile						
5th	533	533	60	37	77	48
10th	535	534	88	67	105	84
25th	539	539	182	147	266	222
50th (median)	544	545	363	322	504	472
75th	552	558	532	530	544	542
90th	585	630	547	546	558	560
95th	685	706	564	559	577	602

Look-2 (interim analysis #2): sample size re-estimation occurred

Table A3-2. Censoring time distribution for time to 1st hospitalization for heart failure in patients who were never hospitalized
[Source: Reviewer's Analysis]

	Look-2 cohort		post Look-2 cohort		Entire population	
In days	Bidil (N=129)	Placebo (N=104)	Bidil (N=304)	Placebo (N=298)	Bidil (N=433)	Placebo (N=402)
Mean	524	539	329	294	388	357
Standard deviation	160	138	181	184	196	204
Range	0-1042	13-1030	5-805	6-701	0-1042	6-1030
Percentile						
5th	90	242	50	28	57	37
10th	451	529	82	58	88	69
25th	538	538	165	119	211	161
50th (median)	541	542	342	292	451	386
75th	549	554	511	462	542	540
90th	572	609	546	545	553	552
95th	672	689	553	550	568	582

Look-2 (interim analysis #2): sample size re-estimation occurred

Appendix 4

Figure A4-1. Log minus log survival plot for all-cause mortality

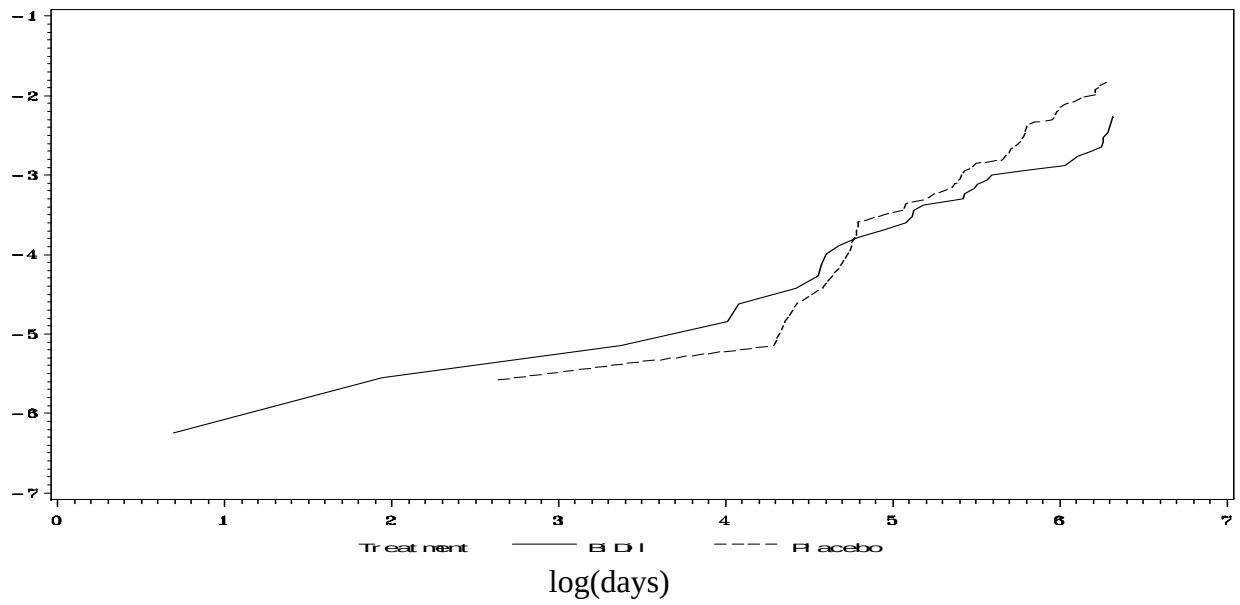


Figure A4-2. Log minus log survival plot for HF hospitalization