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

STATISTICAL REVIEW AND EVALUATION

ADDENDUM #1

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:	May 9, 2005 (E-copy)
DATE OF COMPLETION:	May 10, 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

Review file: c:\data\my ocuments\review\nda\bidil_A-HeFT_1.doc

Table of Contents

1. EXECUTIVE SUMMARY 3

1.1 Conclusions and Recommendations 3
1.2 Brief Overview of Clinical Studies 3
1.3 Statistical Issues and Findings 3

2. INTRODUCTION 4

2.1 Overview 4
2.2 Data Sources 4

3. STATISTICAL EVALUATION 4

3.1 Evaluation of Efficacy 4
3.2 Evaluation of Safety 5

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS 5

4.1 Gender, Race and Age 5
4.2 Other Special/Subgroup Populations 5

5. SUMMARY AND CONCLUSIONS 5

5.1 Statistical Issues and Collective Evidence 5
5.2 Conclusions and Recommendations 6

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

The conclusions in the 5/4/2005 statistical review remain unchanged.

1.2 Brief Overview of Clinical Studies

See the statistical review dated May 4, 2005.

1.3 Statistical Issues and Findings

The primary efficacy endpoint – composite score is a complex endpoint in that two components of the composite score, death at any time on study and first hospitalization for heart failure, are time-dependent. A patient could consequently be included in Interim Analysis #2 (IA2) prior to the occurrence of one or both of these events, and subsequently have either or both events occur prior to the end of the study. This occurred for 32 of the patients enrolled in A-HeFT and included in IA2. The best way for handling the post IA2 data of the IA2 cohort in analysis is a statistical issue needing further research.

The sponsor's 5/10/2005 response correctly pointed out that the reviewer's adjusted analysis ignores the post-IA2 data in these IA2 patients. In the A-HeFT case, this analysis is conservative, giving $p = 0.021$, since in these IA2 patients the bidil group performed better than the placebo group after this interim analysis.

Furthermore, the sponsor explains that this kind of analysis (i.e., ignoring post interim analysis data of the interim analysis cohort) can be anticonservative in some extreme situations where the unadjusted analysis does not achieve statistical significance (Reviewer's note: this point is correct but such inconsistency would have been properly investigated, had it occurred). Thus, the sponsor believes that their adjusted analysis, giving the $p = 0.016$, is the most appropriate analysis that includes post-IA2 data of the IA2 cohort. The sponsor's explanation for why their analysis is most appropriate to include time to event data seems reasonable. However, it is still not clear whether or not mixing the non-time-to-event component with the time-to-event components may complicate the independent increment property required for the Z test for the composite score. This requires further methodological research and thus the validity of their analysis remains questionable at this stage.

In summary, the reviewer's adjusted analysis is preferred for the primary efficacy endpoint in A-HeFT.

2. INTRODUCTION

2.1 Overview

In response to the reviewer's comment dated 4/20/2005 on the adjusted analysis for the primary efficacy endpoint – composite score, the sponsor submitted the 5/10/2005 document (E-copy) to explain why their adjusted analysis is more appropriate than this reviewer's analysis. Adjustment is needed because of the data-driven sample size increase at an interim analysis. This addendum to the statistical review of 5/4/2005 pertains to the sponsor's explanation.

2.2 Data Sources

There are no additional data submitted.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

Results on Primary efficacy endpoint – Composite Score

The primary efficacy endpoint – composite score is a complex endpoint in that two components of the composite score, death at any time on study and first hospitalization for heart failure, are time-dependent. A patient could consequently be included in Interim Analysis #2 (IA2) prior to the occurrence of one or both of these events, and subsequently have either or both events occur prior to the end of the study. This occurred for 32 of the patients enrolled in A-HeFT and included in IA2.

The sponsor's 5/10/2005 response correctly pointed out that the reviewer's adjusted analysis ignores the post-IA2 data in these IA2 patients. In the A-HeFT case, this analysis is conservative, giving p = 0.021 (Table 4 of the 5/4/2005 statistical review, also shown below), since in these IA2 patients the Bidil group performed better than the placebo group after this interim analysis.

Furthermore, the sponsor explains that this kind of analysis (i.e., ignoring post interim analysis data of the interim analysis cohort) can be anticonservative in some extreme situations where the unadjusted analysis does not achieve statistical significance (reviewer's note: the sponsor's point is correct but such inconsistency would have been properly investigated, had it occurred). Thus, the sponsor believes that their adjusted analysis, giving the p = 0.016 (also in the table below), is the most appropriate analysis that includes post-IA2 data of the IA2 cohort. The sponsor's explanation for why their analysis is most appropriate to include time to event data seems reasonable. However, it is still not clear whether or not mixing the non-time-to-event component with the time-to-event components may complicate the independent increment property required for the Z test for the composite score. This requires further methodological research and thus the validity of the sponsor's analysis remains questionable at this stage.

In summary, the reviewer's adjusted analysis is preferred for the primary efficacy endpoint in A-HeFT.

Table 4a. Mean composite score of death, first hospitalization for heart failure, and change from baseline in six-month quality of life – primary efficacy endpoint
[*Source: Table 4 of 5/4/2005 review; 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

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

See the 5/4/2005 statistical review.

4.2 Other Special/Subgroup Populations

See the 5/4/2005 statistical review.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

The primary efficacy endpoint – composite score is a complex endpoint in that two components of the composite score, death at any time on study and first hospitalization for heart failure, are time-dependent. A patient could consequently be included in Interim Analysis #2 (IA2) prior to the occurrence of one or both of these events, and subsequently have either or both events occur prior to the end of the study. This occurred for 32 of the patients enrolled in A-HeFT and included in IA2. The best way for handling the post IA2 data of the IA2 cohort in analysis is a statistical issue needing further research.

The sponsor's 5/10/2005 response correctly pointed out that the reviewer's adjusted analysis ignores the post-IA2 data in these IA2 patients. In the A-HeFT case, this analysis is conservative, giving p = 0.021 (see Table 4a), since in these IA2 patients the bidil group performed better than the placebo group after this interim analysis.

Furthermore, the sponsor explains that this kind of analysis (i.e., ignoring post interim analysis data of the interim analysis cohort) can be anticonservative in some extreme situations where the unadjusted analysis does not achieve statistical significance (Reviewer's note: this point is correct but such inconsistency would have been properly investigated, had it occurred). Thus, the sponsor believes that their adjusted analysis, giving the p = 0.016 (see Table 4a), is the most appropriate analysis that includes post-IA2 data of the IA2 cohort. The sponsor's explanation for why their analysis is most appropriate to include time to event data seems reasonable. However, it is not clear whether or not mixing the non-time-to-event component with the time-to-event components may complicate the independent increment property required for the Z test for the composite score. This requires further methodological research and thus the validity of their analysis remains questionable at this stage.

In summary, the reviewer's adjusted analysis is preferred for the primary efficacy endpoint in A-HeFT.

5.2 Conclusions and Recommendations

The conclusions in the 5/4/2005 statistical review remain unchanged.