DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

CARDIOVASCULAR AND RENAL DRUGS

ADVISORY COMMITTEE

VOLUME II

Thursday, June 16, 2005

8:00 a.m.

Holiday Inn Gaithersburg
Two Montgomery Avenue
Gaithersburg, Maryland

PARTICIPANTS

Steven E. Nissen, MD, FACC, Chair
LT. Cathy Groupe, RN, BSN, Executive Secretary

MEMBERS:

Susanna L. Cunningham, PhD, Consumer Representative
William R. Hiatt, MD
Frederick J. Kaskel, MD, PhD
Ronald Portman, MD
John R. Teerlink, MD

CONSULTANTS (VOTING):

Jonathan Sackner-Bernstein, MD
Robert J. Samuels
Thomas Fleming, PhD
Vivian Ota Wang, PhD

CONSULTANT (NON-VOTING):

Charles H. McCleskey, MD

FDA STAFF:

Norman Stockbridge, MD, PhD
Robert Temple, MD
Douglas Throckmorton, MD

C O N T E N T S

C O N T E N T S (Continued)

P R O C E E D I N G S

Call to Order

DR. NISSEN: If people will take their seats, I would like to get started. We have lots of work to do today. Let's begin with some introductions and we will start over here. Dr. McCleskey, please tell us who you are, what you represent and where you are from.

DR. MCCLESKEY: My name is Dr. Charles McCleskey. I am an anesthesiologist. I work for Abbott Laboratories and I am an interim industry representative on this committee. I normally sit on the Anesthesia and Life Support Committee.

DR. OTA WANG: Good morning. I am Vivian Ota Wang. I am a geneticist and behavioral scientist. I am from the Ethical, Legal and Social Implications Program of the National Human Genome Research Institute at NIH.

DR. CUNNINGHAM: Good morning. My name is Susanna Cunningham. I am a professor at the University of Washington School of Nursing and I am on the committee as the consumer representative.

DR. HIATT: I am William Hiatt, University of Colorado, vascular medicine.

DR. PORTMAN: Ron Portman, pediatric nephrology and hypertension, University of Texas-Houston.

DR. KASKEL: Rick Kaskel, Albert Einstein College of Medicine, pediatrics and nephrology.

DR. NISSEN: I am Steve Nissen. I am a cardiologist from the Cleveland Clinic.

LT. GROUPE: LT. Cathy Groupe. I am the executive secretary for the Cardiovascular and Renal Drugs Advisory Committee.

DR. TEERLINK: John Teerlink, University of California San Francisco and San Francisco VA Medical Center, heart failure.

DR. SACKNER-BERNSTEIN: Jonathan Sackner-Bernstein, cardiologist from St. Luke's-Roosevelt Hospital Center.

DR. FLEMING: Thomas Fleming, Department of Biostatistics, University of Washington.

DR. STOCKBRIDGE: I am Norman Stockbridge. I am the Acting Director of the Division of

Cardiorenal Drug Products.

DR. THROCKMORTON: Doug Throckmorton. I am the Acting Deputy Center Director, on leave from the Cardiorenal Division.

MR. SAMUELS: Good morning. My name is Bob Samuels and I am the patient representative on the committee.

DR. NISSEN: I think we have one empty seat for Dr. Bob Temple from the FDA who, I am certain, will be here. Cathy, I think you are going to do the conflict of interest statement so, please, proceed.

Conflict of Interest Statement

LT. GROUPE: The following announcement addresses the issue of conflict of interest and is made a part of the record to preclude even the appearance of such at this meeting:

Based on the submitted agenda and all financial interests reported by the committee participants, it has been determined that all interests in firms regulated by the Center for Drug Evaluation and Research present no potential for an

appearance of a conflict of interest with the following exceptions: In accordance with 18 U.S.C. 208(b)(3), full waivers have been granted to the following participants, Dr. Steven Nissen for consulting for a competitor for which he receives less than $10,001 per year.

A copy of the waiver statements may be obtained by submitting a written request to the agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

In addition, we would like to acknowledge for the record that Dr. David DeMets will be participating on behalf of the sponsor, NitroMed, with the stipulation that he was recused from the June 15th and 16th, 2005 Cardiovascular and Renal Drugs Advisory Committee meetings.

With respect to FDA's invited industry representative, we would like to disclose that Dr. Charles McCleskey is participating in this meeting as an acting industry representative, acting on behalf of regulated industry. Dr. McCleskey is employed by Abbott Laboratories.

In the event that the discussions involve any other products for firms not already on the agenda for which an FDA participants has a financial interest, the participants are aware of the need to exclude themselves from such involvement and their exclusion will be noted for the record. With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose products they may wish to comment upon.

DR. NISSEN: Norman, you wanted to make some introductory comments.

DR. STOCKBRIDGE: No, actually I don't. I am happy to welcome everybody who is participating in this meeting today, and appreciate your service but I have no comments at all to make about the topic. Thank you.

DR. NISSEN: Then we will stay ahead of schedule, which won't last for very long by my experience with these committees, but we will try. Let's get into the sponsor presentation. So, I

think we will turn the floor over to the sponsor and we would like to hear what you have to say.

Sponsor Presentation

Background and Introductions

DR. WORCEL: Good morning, ladies and gentlemen. Dr. Nissen, members of the advisory committee, Drs. Temple, Stockbridge and Throckmorton, officers and reviewers of the FDA, ladies and gentlemen, my name is Manuel Worcel and I am the chief medical officer at NitroMed. On behalf of NitroMed, I would like to thank you for the opportunity to review the evidence supporting the approval of BiDil.

BiDil is a fixed dose combination tablet containing the two active drugs, isosorbide dinitrate and hydralazine hydrochloride. Each tablet contains 20 mg of isosorbide dinitrate and 37.5 mg of hydralazine hydrochloride. Isosorbide dinitrate is a vasodilator of the large and small arteries and at therapeutic doses is a preferential venous dilator. Its dilator properties result from release of nitric oxide and the subsequent

activation of guanylyl cyclase. Hydralazine is a selective dilator of small artery smooth muscle.

Today we will present the findings of three large trials of isosorbide dinitrate and hydralazine. The first vasodilator heart failure trial, V-HeFT I, was conducted from 1980 to 1985. This trial compared the ISDN/hydralazine combination to placebo and another vasodilator, prazosin. It was run exclusively at VA hospitals and included 642 white and black men.

The second vasodilator heart failure trial, V-HeFT II, was a trial comparing the effects of the ISDN/hydralazine combination to the effects of enalapril in 804 men, and was also conducted in VA hospitals.

Post hoc analyses from these two trials led to the design of a study of the fixed dose combination in black heart failure patients, the African American Heart Failure Trial or A-HeFT. In A-HeFT 1050 black men and women were randomized to receive standard heart failure therapy plus BiDil or to receive standard heart failure therapy plus

placebo. A-HeFT was conducted between 2001 and 2004.

This slide summarizes key milestones in the regulatory history of BiDil. The initial application was submitted in the mid-1990s by Medco Research and was based on the results of the B-HeFT trials analyzing data from the overall cohorts including white and black patients.

Following presentation to the Cardiorenal Advisory Committee meeting in February, 1997, the FDA issued a "non-approvable" letter. During the 1990s it became known that the responses to dosing converting enzymes on beta-blockers in black hypertensive patients were lower than in white patients.

Following this evidence, Dr. Carson et al. reevaluated the effect of enalapril and ISDN/hydralazine in V-HeFT I and II. This retrospective analysis showed that BiDil appeared to have a greater effect on survival and other clinical endpoints in black heart failure patients. This post hoc reanalysis generated the hypothesis

that ISDN/hydralazine would particularly benefit black patients.

We approached the FDA with a reanalysis in 1999 and 2000 and together developed a clinical plan to confirm the response to BiDil in black patients with heart failure. As a result of the discussions, the FDA informed NitroMed in 2001 that an additional clearly positive trial in African Americans would form the basis for approval of BiDil in black patients.

A-HeFT was then started in June, 2001, and in July, 2004 the trial was terminated early based on the recommendations of the data and safety monitoring board. The board found a favorable mortality benefit in patients receiving BiDil. Then an amendment was submitted on December 23, 2004.

The results of the clinical development program for BiDil in black patients with heart failure demonstrate that BiDil produces a meaningful reduction in the risk of mortality; a meaningful reduction in the risk of heart failure

hospitalizations; and a meaningful improvement in quality of life. BiDil was also generally safe and well tolerated, including in patients with a wide range of symptoms and background medications.

Based on these findings, we have proposed the following indication for BiDil: BiDil is indicated for the treatment of heart failure as an adjunct to standard therapy in black patients to improve survival, prolong time to hospitalization for heart failure, and improve quality of life.

This is our program for today. Following my introduction, Dr. Jay Cohn, from the University of Minnesota Medical School, will review key results of the V-HeFT I and V-HeFT II trials showing the signal that we followed in the design of the African American heart failure trial. Dr. Anne Taylor, chairperson of the A-HeFT steering committee, also at the University of Minnesota, will then outline the design of the A-HeFT study in black patients. Following Dr. Taylor, Dr. Yancy, of the University of Texas Southwestern Medical Center, will present the A-HeFT results. I should

say that Dr. Cohn, Dr. Taylor and Dr. Yancy are members of the A-HeFT steering committee. Dr. Milton Packer, also of the University of Texas Southwestern Medical Center, will conclude with an integrated summary of the findings. At the end, we will open the floor to a question and answer period which will be moderated by Dr. Sabolinski.

Throughout today's program, we will be assisted by several colleagues from NitroMed who will provide scientific support. In addition, the advisors listed on this slide are present to potentially answer questions.

Now I would like to turn the microphone over to Dr. Cohn.

V-HeFT I and V-HeFT II: ISDN/HYD Effects in Black Patients

DR. COHN: Thank you, Manuel and good morning to all of you. As Dr. Worcel has pointed out, we are going to be discussing today three trials, briefly the V-HeFT I and V-HeFT II trials which utilized generic isosorbide dinitrate and hydralazine and then, in greater detail, the A-HeFT

trial which used the fixed dose combination now called BiDil.

V-HEFT I and V-HeFT II were carried out in VA medical centers in the 1980s. At the time they were designed they were, in fact, the first trials to be carried out to study drug effects in chronic heart failure. They were supported by the VA cooperative studies program so they were not designed specifically for the drug approval process, but they were to answer a scientific question, that is, does vasodilator therapy favorably affect the course of chronic heart failure? At the time these trials were designed the only therapy for heart failure was digitalis and diuretics so that was background therapy. There were no other drugs that had been developed at that point to influence the course of heart failure.

So, we selected a patient population in the VA medical centers and, since the majority of such patients are males, we decided to exclude females from this trial and try to maintain

homogeneity in the patient population. So, these were men aged 18-75 years old. They had all had what was felt clinically to be symptomatic heart failure for more than three months.

We did bicycle ergometry exercise testing in all patients with gas exchange measurements so we could measure maximum oxygen consumption during exercise, and the entrance criteria was, in fact, a peak oxygen consumption of less than 25 mL/kg/min so this was documented impairment of exercise tolerance. The patients were symptomatic despite digitalis and diuretic therapy and they had objective evidence of cardiac enlargement, either an enlarged heart on chest x-ray, a reduced ejection fraction by whatever method was used at that center, and that had to be less than 45 percent, and a dilated chamber, measured by echocardiography with an end diastolic dimension, transverse diameter of greater than 2.7 cm/m2 body surface area. These were the criteria for entrance in both V-HeFT I and V-HeFT II.

I point out the exclusion criteria because

they were a little different than they would have been today. We excluded patients who had hypertension that required therapy other than diuretics. We didn't want to have other vasodilators on board. We excluded angina that required frequent or chronic nitrate therapy since we were administering isosorbide dinitrate. And we excluded patients who were on beta-blockers or other vasodilators because in the 1980s beta-blockers were contraindicated in the treatment of heart failure. So, you can see how far we have come in the last couple of decades. We excluded patients who had myocardial infarction or cardiac surgery within the previous three months, and we excluded people with hypertrophic cardiomyopathy or valvular heart disease or severe other comorbidities which would limit life expectancy. So, those were the entrance criteria in both trials.

V-HeFT I was designed with three arms because we were evaluating the efficacy of vasodilator therapy and there had been two

vasodilator regimens at that time that had been demonstrated to exhibit a favorable effect on hemodynamics and on pump function. Since we had had no experience with long-term follow-up in patients receiving dig. and diuretic, and we wanted the power to actually combine these two groups if, in fact, they had a similar response, we randomized more patients to the placebo arm than to either of the vasodilator arms to attain maximum power.

There were 276 patients placed on placebo, double-blind, and 183 who were given prazosin, the alpha-blocker, and the dose of prazosin was 5 mg 4 times daily. There were 186 patients given the generic form of ISD in a target dose of 40 mg 4 times daily and hydralazine in a target dose of 75 mg 4 times daily. All of these drugs were dummy controlled so that it was all double blinded.

The follow-up in V-HeFT I was to a maximum of 5.7 years. We were slow recruiting. We had a very limited number of VA medical centers in that trial so that it took time and from beginning to end of the trial the maximum follow-up was 5.7

years. The mean follow-up by protocol was 6 months and the mean follow-up was 2.3 years.

When we completed V-HeFT I the steering committee felt that we had demonstrated a significant benefit of the ISDN/H treatment arm. I will abbreviate that in the future as I/H for simplicity.

As a result of that, which was viewed as a significant benefit, the steering committee felt it was unethical to include a placebo arm in the follow-up study, V-HeFT II. So, rather than a placebo arm, we had a two-arm trial in which we compared the winning agent in V-HeFT I, I/H, with a newer form of therapy which had been studied in small trials during the time we were doing V-HeFT I, and that was the converting enzyme inhibitors, and we chose enalapril, and we gave enalapril at a dose of 10 mg twice daily and. Once again, this was double blinded and there were 400 patients in each treatment arm. The follow-up again was about 5 years maximum and the mean follow-up was 2.5 years.

Since we were asking a scientific question as to whether vasodilator therapy would influence the course of the disease, we looked at a number of endpoints and we called them major endpoints at that time. This was in 1979 when we designed this trial. All-cause mortality was clearly the most important endpoint and we powered the trial based upon prediction of what could happen. We had no data on mortality in this population in a controlled environment so we had to make some guesses.

In addition, since we felt this would be a sick population and they would probably die at a rapid rate, we felt that looking at a single time point might give us a better discriminator between these two therapies so we chose a two-year all-cause mortality as another primary endpoint. And the mechanisms of death were adjudicated.

We were interested in hospitalizations at that time and we felt that was an important endpoint so as a primary endpoint we said we would look at the number and duration of cardiovascular

hospitalizations. These were not adjudicated in this early trial. We accepted the investigator designation of the reason for hospitalization.

We did sequential exercise testing with maximum oxygen consumption at regular intervals during the trial to quantitate the improvement, if there was any, in exercise performance. In V-HeFT II, for the first time, we initiated a quality of life assessment which we had developed specifically for this protocol.

Well, this was the survival curve in V-HeFT I. This is the curve that led the steering committee to feel that we had demonstrated a favorable effect of the I/H combination, in yellow, compared to the placebo, in blue, and the prazosin, in green, which tracked together on this inexorable downhill course with an annual mortality rate of 20 percent in the placebo arm and the prazosin arm, and a mortality rate of 12 percent in the ISDN/HYD arm. The log-rank p value for that difference, which is the standard approach that the FDA asked us to take, was 0.093.

Now, when that trial was terminated, the investigators had used a Cox model to adjust for baseline differences, and using the Cox model the p value was 0.046. So, there was some discussion about what the true p value was but, in fact, this did not achieve the traditional level of statistical significance.

These are the more precise data on the overall mortality in V-HeFT I. The I/H arm was compared to the placebo arm and there were 44 percent of the placebo arm who died during the trial and 38.7 percent of the I/H arm with prazosin. Of course, it is 44 percent with placebo and even a little higher with prazosin, and the risk ratio for prazosin was 1.11; for ISDN/HYD, 0.78 and that was the p value of 0.093.

The two-year endpoint is down on the bottom. There was at two years a 34.3 percent mortality in the placebo group and a 25.6 percent mortality in the I/H group, and that p value was 0.053. I want you to remember that 25.6 value in V-HeFT I because that is the I/H arm which we

attempted to replicate in V-HeFT II in the same population.

These were the V-HeFT II survival curves. Of course, both treatments exhibited a rather parallel decline in survival with, at all time points, slightly better survival in the enalapril arm than the I/H arm. The overall p value for this difference was 0.083 and, remember, there were 800 patients in this trial. We are still talking about small studies compared to current trial design. And that p value was 0.083, exhibiting a trend, certainly, with a hazard ratio of 1.23 for a better outcome in the enalapril-treated group.

This is, in fact, the data on V-HeFT II in detail. There were 132 deaths on enalapril, 153 on I/H. That is the 1.23 hazard ratio and a log-rank of 0.083. Now, at a 2-year time point. The mortality in the enalapril group was 18 percent, and look at the I/H arm, 25 percent. As you remember from the previous slide, it was 25.6 percent in V-HeFT I. So, we felt that this was remarkable replication in the same centers, with

the same patient population, with the same background therapy of dig. and diuretic and, in fact, we felt it was almost justified to place the placebo arm in this context. Remember, that was a 34.3 percent mortality. The p value for this 2-year time point was 0.016.

This implied that ACE inhibitors were preferential therapy and, since ACE inhibitors were in fact a new chemical entities that were heavily marketed, they became standard of therapy, and over the subsequent years the use of ISDN/HYD became confined to those cognizenti who were aware of the data and who had participated or professed the benefits of this form of therapy.

Over the ensuing years it became apparent that enalapril and ACE inhibitors exhibited lesser benefit in hypertension in black patients than in white patients. In fact, that difference has become part of labeling for ACE inhibitors in the treatment of hypertension.

Armed with that growing evidence that there must be a differential response, for many

reasons that we need not go into, we went back to the V-HeFT II data and reanalyzed it based on self-designated race. The reason for doing that was entirely because of the clear evidence that in hypertension there was a differential response.

This was the result of our reanalysis of V-HeFT II. When we looked at the black patients who identified themselves as black in V-HeFT II, the survival in the enalapril and the I/H arms was identical, superimposing the hazard ratio of 1.01. There were only 215 of these patients who called themselves black so it is a small population, of course, and very little power. There were 574 white people in V-HeFT II, and in that group there was a striking benefit of enalapril compared to I/H. The p value in this small population was 0.002 and the hazard ratio was 1.39.

So, there were two possible explanations for this. If in fact this was true, and it appeared to be, either ACE inhibitors are less effective in black people, for which there was strong evidence, and/or I/H is more effective in

black people and one could not, from these data, separate those two possibilities.

So, we went back to V-HeFT I and analyzed the racial difference in response and here are the data. Among the 128 black patients in V-HeFT I the benefit of H/I compared to placebo was striking and the p value for that difference was 0.04 and hazard ratio of 0.53 suggesting a 47 percent risk reduction.

In contrast, in the white patients in V-HeFT I, and there were 324 of those, the curves were much closer together. The hazard ratio was 0.88. There is, in fact, a nominal 12 percent reduction in mortality risk but, of course, the p value was nowhere near significant. So, we felt these data certainly provided support for the idea that there may be, in fact, not only a differential response to ACE inhibitors but a differential response to isosorbide dinitrate and hydralazine.

So, we could conclude from this analysis that ISDN/hydralazine in V-HeFT I compared to placebo was associated with a 22 percent lower risk

of death overall and that was not quite statistically significant; a 12 percent lower risk of death in white patients which, of course, was not significant; and a 47 percent lower risk of death in black patients, with a p of 0.04.

In V-HeFT II we could conclude that enalapril compared to ISDN/hydralazine was associated with a 23 percent lower mortality overall, once again not significant, but a 39 percent lower mortality in white patients, and that was significant, and no difference in mortality in black patients.

Now, we presented these data to the FDA and they agreed that our hypothesis was, in fact, attractive. As Dr. Worcel has pointed out, they informed us that a confirmatory study in black patients could be the basis for approval of the drug in that population.

Based on V-HeFT I and V-HeFT II, we agreed with the agency that a clinical study was needed to confirm the hypothesis that isosorbide dinitrate and hydralazine combination benefits outcomes in

black patients. We, therefore, designed A-HeFT as a prospective, placebo-controlled study with the objective of testing BiDil's effects on survival, heart failure hospitalizations, and quality of life now in patients receiving contemporary therapy for heart failure. Remember, this is a whole new era. The original studies with this drug combination were in patients receiving only dig. and diuretic. We were now embarking on a study in patients receiving ACE inhibitors, angiotensin receptor blockers, beta-blockers, aldosterone antagonists, and Dr. Taylor will describe to you how well treated this population was.

So, the burden was in fact considerably higher and the mortality reduction that we had identified in our retrospective analysis of V-HeFT I, which was a 47 percent mortality reduction--we were optimistic that we would be able to confirm that but we didn't expect necessarily that the benefit would be as obvious in patients who were already so well treated with drugs, all of which have been shown to prolong life. So, I will now

turn the podium over to Dr. Taylor who will describe the protocol that we used in A-HeFT.

Questions from the Committee

DR. NISSEN: Before we do that, I think this might be an opportunity for some questions for Dr. Cohn, if it pleases the committee. Anybody want to ask any detailed questions? Is everybody satisfied? I will have some for you later but let me just ask one question. Obviously, the question is the mechanism here and the suggestion that there may be some gene responsible for this. Is there anything you can tell us about this that explains the differential response?

DR. COHN: Well, the working hypothesis has been that there is evidence for reduced nitric oxide bioactivity in African American populations on average compared to white populations, and that data has been generated in a number of laboratories over the last decade using methods for studying nitric oxide stimulating substances or mechanisms. And it does appear that black people, for reasons which we certainly do not know, exhibit on average

a less robust response to this released nitric oxide. In fact, that provides the physiologic underpinnings for why we might have expected some differential response to an agent, BiDil, which is a nitric oxide enhancing therapy, and we believe that its action is mediated by nitric oxide which is released by the ISDN and preserved by the antioxidant properties of hydralazine. So, there are very good biological underpinnings to this differential response. But the identification of the differential response really came from our mortality evaluation.

DR. HIATT: The data you presented would be considered Phase II I think in terms of the level of evidence, and the usual regulatory hurdle would be two pivotal Phase III trials. Can you just give us a little bit of background on why one was required by the FDA, not two?

DR. COHN: Well, I think you might want to ask the FDA about their view. In the original presentation of isosorbide dinitrate/hydralazine the FDA had accepted the improvement of exercise

tolerance, which I have not shown you today, that was demonstrated in both V-HeFT I and in V-HeFT II. They accepted that as one pivotal study for efficacy of this combination but that was, of course, in the overall population. We have very little power in the subgroup analysis to look at these other endpoints. The agency did, in fact, claim though that another outcome trial would be adequate for registration.

DR. NISSEN: Tom Fleming?

DR. FLEMING: I think that is a key issue and my sense or interpretation would be similar to yours that the V-HeFT trials essentially are hypothesis generating here in the way that they were designed and in the way that they were analyzed. Hence, I would look for what I call strength of evidence of two trials in a confirmatory trial, i.e., these results are going to have to be highly statistically persuasive because there is just a single trial.

But, Jay, kind of leading up to what we should expect, A-HeFT, A-HeFT has targeted a

composite endpoint and principal components of that composite were at least in part studied in the V-HeFT I and II trials. Among those are mortality and time to first hospitalization, and I think we will probably have a fair amount of discussion about those endpoints in A-HeFT. You have reviewed with us the mortality results. Can you go through with us, with some care here, the time to first hospitalization results in V-HeFT I and V-HeFT II since they were actually among the array of your "major" endpoints?

DR. COHN: We hadn't intended to do that because it gets to be a very complex assessment. I think you will see in the final conclusions some data on hospitalization in V-HeFT I and V-HeFT II. The trends were all in the right direction.

DR. FLEMING: Do you have slides, Jay?

DR. COHN: I do not have the slides--yes, here we have a slide of the time to first heart failure hospitalization for all patients in V-HeFT I and in V-HeFT II. As you can see, there was a clear trend for V-HeFT I to have a delay in first

hospitalization compared to placebo, in blue. Here, in V-HeFT II you see that they are superimposed. We know from other trials that ACE inhibitors do delay hospitalization for heart failure so this would support the idea that H/I not only worked against placebo but was not inferior to V-HeFT II but, of course, there is very little power here.

DR. FLEMING: You show us these results by race. In fact, I think I am looking at them in your briefing document.

DR. COHN: Yes, this is the breakdown by race. These are the black patients in V-HeFT I and V-HeFT II once again showing this trend to the curve for ISDN/hydralazine to be more favorable than placebo. In the black population it also appears to be more favorable than enalapril. Once again, the statistics aren't shown here. This is a very small sample size but the trend is in the right direction.

DR. FLEMING: And then in the non-black patients?

DR. COHN: I am not sure--do we have that on a slide in the non-blacks?

Well, in fact this separates the V-HeFT I into black and white and you can make as you will of those curves for first heart failure hospitalization.

Do we have that in V-HeFT II as well? Here is the V-HeFT II data in black and white patients, all consistent with the hypothesis that there appears to be a little better outcome in the black patients on ISDN compared to enalapril and no difference in the white patients.

DR. FLEMING: It might actually be easier to look, at least for those of us on the advisory committee, in your briefing document because the curves you are showing don't seem to be exactly the same as what you have shown us in the briefing document. I am looking specifically on page 30, 31 and 32. So, in V-HeFT I for time to first hospitalization for heart failure the curves are showing slight separation at a year and then overlapping overall.

Then, when we look on 31 and 32, breaking it out into African American versus white, there is no indication of interaction there--

DR. COHN: These are first heart failure hospitalizations. Make sure you are looking at the same category.

DR. FLEMING: I am looking at time to first hospitalization for CHF.

DR. COHN: Maybe Dr. Sabolinski can address this point.

DR. SABOLINSKI: Thank you. Mike Sabolinski, NitroMed. The curves that Dr. Cohn has shown are truncated and the ones that are in our briefing book are not.

DR. FLEMING: So, I think then the briefing book is giving us a bit more comprehensive evidence.

DR. SABOLINSKI: What you are seeing in the briefing book is a pattern that over two years, when there is a meaningful number of patients, the curves do separate and they do separate primarily by race. Beyond two years there really is not a

meaningful number of patients to draw conclusions.

DR. FLEMING: Actually, there appears to be a considerable amount of evidence out there. In essence, if we look in the briefing documents for the advisory committee on pages 30, 31 and 32 it looks as though there are some short-term emerging differences that do not persist so that over time the overall distributions of time to first hospitalization are relatively comparable, and not much evidence of interaction by race.

Then if we look at V-HeFT II, on pages 52, 53 and 54 we find that the curves are overlapping over the first two years, as Jay was showing, and then there is some separation favoring enalapril. If we look by interaction, modest interaction--I would call it a very modest suggestion of interaction with slightly--slightly--better results in African Americans; slightly worse results in whites.

DR. COHN: Yes, I think there are a couple of issues and I think Dr. Packer wants to make a comment too, but let me remind you of a couple of

important issues.

The hospitalization data, of course, is impacted by a 20 percent annual mortality which is high, of course, and the differential mortality--

DR. FLEMING: A valid point, so is it as we would want it to be for hospitalization-free survival?

DR. COHN: Those curves I think are just hospitalization, unfortunately.

DR. FLEMING: So, you are censoring the deaths?

DR. COHN: So, the deaths were in fact censored for that analysis. The other issue is when you see the A-HeFT data you must keep in mind that in the early 1980s we had no disease management strategies and patients were hospitalized frequently for worsening heart failure because we had not yet learned how to keep them out of the hospital, which we now do far more effectively. So, the experience in hospitalizations in the 1980s is not easy to extrapolate to the 21st century, and that will

potentially impact a bit on how you interpret these data, but I fully agree that--

DR. FLEMING: That is a valid point. In fact, it is part of the reason that some of us would indicate that these are Phase II trials. The context in which these studies were done differs from today's context and so--

DR. TEMPLE: Death is similar.

DR. FLEMING: Pardon me?

DR. TEMPLE: Death is similar.

DR. FLEMING: Death is similar but Jay makes the point that death or hospitalization might have been something that would have occurred in a different context back in that era. There are different ways of managing patients. Of course, back then we didn't have the ACE inhibitors and the beta-blockers and calcium channel blockers, and all. I guess the point is that while you can validly say one has to have some caution in interpreting the results because of that, that look relatively unfavorable. Anything that you might state that looks a little bit favorable or

suggestive of positive results equally has to be viewed in that context.

So, in essence, at least my interpretation here is that the trials were designed to look at mortality and time to first heart failure hospitalization, and they provided evidence on both. Both of them have to be taken with caution because of the difference in the era. But the first hospitalization results suggest to me that there is much more modest evidence of interaction by race. There are relatively modest effects. One should not just look at the first one year or two years. The results over a longer time frame are less impressive, I would argue.

DR. COHN: Yes, but remember, the differential mortality was leaving many more patients at risk for hospitalization in the ISDN/hydralazine arm in V-HeFT I.

DR. FLEMING: That shouldn't affect V-HeFT II.

DR. COHN: No, not as much.

DR. FLEMING: That is not a huge

difference.

DR. COHN: Tom, I must tell you also that my confidence in the hospitalization data in V-HeFT has greatly impacted on our more recent experience that adjudication of hospitalizations makes an exceedingly big difference in the interpretation. These data were collected in multiple centers by investigator adjudication, not by a central process. So, I must say, I don't have a lot of confidence in those data. We showed them to you for completeness but without attempting to claim that they are very valuable.

DR. PACKER: Tom, it is actually more complicated than that. The way the hospitalization data, aside from the issues that you have already heard, which is lack of adjudication--in spite of the issues of competing risk with mortality, there is a third confounding factor and that is that in the current era when hospitalization occurs in a clinical trial it is recorded immediately upon its occurrence. In the V-HeFT trial the hospitalization data weren't collected at the time

of occurrence. They were collected at the scheduled visit that followed the event, which means that a hospitalization could have occurred in January and not have been recorded until April. There were no dates of hospitalization, the actual occurrence of hospitalization in V-HeFT.

So, we have an issue here in terms of when you do a time-to-event analysis in terms of trying to identify with precision the actual occurrence of the hospitalization.

DR. FLEMING: That is a valid point that typically could lead to a modest to slight attenuation.

DR. PACKER: It depends on the magnitude of the treatment difference and the time shift and it is really hard.

DR. NISSEN: You know, I am old enough, Tom, to have practiced in the dark ages of medicine, back in the 1980s and I can tell you that clinical trials--you know, the rigor that we now understand didn't exist. These were some of the first clinical trials in cardiology of this size

ever conducted, and the size, of course, is very modest. I think it is also important that all of us recognize that this is a post hoc analysis. So, all the limitations that are present in trials done in an earlier era and with a post hoc analysis have to be taken into account, and I think you are pointing out some of those limitations.

DR. FLEMING: In fact, it is clearly another aspect of why there are concerns about how you would interpret this. But the real post hoc aspect of this was the race. It wasn't post hoc to look at hospitalization, and Jay was telling us it was one of their major endpoints.

DR. NISSEN: Yes, I think that is correct. Jonathan?

DR. SACKNER-BERNSTEIN: In terms of hospitalization, I am wondering if you could share with us the total number of hospitalizations in the black patients in the two V-HeFT studies; total number, if you have it, of days in house for heart failure hospitalizations; total heart failure hospitalizations between the groups, with all the

caveats we have discussed.

DR. COHN: I don't think we have those at all and they are not in your briefing document either because they weren't collected. We admit weakness on the hospitalization issue. Given that this was planned in 1979, we weren't perceptive about how important that was going to be.

DR. SACKNER-BERNSTEIN: One other point, getting back to the Phase II kind of flavor to the data, perhaps the FDA could confirm something, if it is relevant, from the briefing document, on page 17, where it talks about the manner in which--at the bottom of the page--we are to use the data from V-HeFT I and V-HeFT II in this process. Basically it says, a concern regarding the bioequivalence of the formulations between V-HeFT and A-HeFT was raised at a meting in November of '92. Then it says therefore, the post hoc analysis results of efficacy in the two trials will not be used for support of efficacy.

I interpret that to be an FDA confirmation of the concern that you reiterated about the

utility of the two V-HeFT studies. Is that a correct interpretation?

DR. TEMPLE: If it says because of the uncertainty about whether the products are identical we won't look at it, I totally disagree. Somebody may have said that but I don't agree with that. We don't think it is not there, or anything. How much to make of it is a separate question for all the reasons you have given but I don't believe it is because we have some doubts about the product.

DR. HIATT: From the FDA's point of view, what level of evidence do we need today based now on data clearly being from a different time and supportive?

DR. NISSEN: It is what you think the level of evidence you need is. I mean, that is why you are here, to provide clinical judgment on whether the level of evidence meets our standards for robustness that we want for a company to get approval. So, that is why we are here, to answer that question.

DR. TEMPLE: I mean, if you want some sense of history, the first outcome data that ever got into anything in heart failure was based on one outcome study consensus, along with evidence of improved function in one of two. By the way, the U.S. study failed completely; the foreign study won for enalapril. Subsequent individual claims in various subsets of the population, class II and so on, have each been supported by one study. However, the results were often very robust consensus. One had a p value out to four zeroes. So, we have generally thought that outcome data should be supported by either two trials, which is very hard to do if you have one in hand, or a single trial with a robust finding.

DR. NISSEN: I might add that on this committee we have occasionally viewed mortality somewhat differently from morbidity. I remember a discussion about a trial where we talked about the implied mortality endpoint and how to view that. Tom, I think you were there as well. So, I think that is important as we factor in these

discussions.

But, Jay, I don't think what anybody is saying here is inconsistent. You are saying that you consider these analyses to be hypothesis generating and I think several members of the committee are suggesting the same concept.

DR. FLEMING: I think you are referring to January 7, 2003 carvedolol discussion--

DR. NISSEN: Yes. Only Tom Fleming remembers the actual dates--

[Laughter]

--can you tell me the time of the day that we actually had the discussion?

DR. FLEMING: It was mid-morning. Bob Temple was there.

DR. NISSEN: That is just sick, Tom!

DR. FLEMING: In essence though, just to follow-up on Jonathan's point, at least in the briefing document it doesn't say the reasons why, although I think everything that is on the table is part of the justification. It says that for efficacy A-HeFT was the only source for the review

but for safety additional data from V-HeFT was used. I think, for reasons that we can get into later one, the nature of the post hoc analyses by race, etc., are the type of analyses that in most cases, even if there was more relevance to how the V-HeFT trials were done relative to the modern era, you would still look at as hypothesis generating. But for all the reasons that have been put forward in terms of the supportive care, lack of ACE inhibitors, calcium channel blockers, beta-blockers and what you are now telling us today, your concerns about relevance of how hospitalization was managed, all lead to the observation that these results are certainly of some insight but have to be viewed with real caution.

DR. TEERLINK: Vis-a-vis how to interpret the V-HeFT, particularly the mortality data, I was intrigued to see that the interaction effect is not significant. It is an interaction effect of 0.15 looking at race in terms of mortality. So, being a statistical neophyte, I always thought that we first looked at the interaction effect and if the

interaction effect was positive, then we looked at the individual aspects of that. But if the interaction effect was negative, we kind of said, well, that may be interesting but certainly doesn't provide any statistical support.

DR. COHN: Well, your observation is in part right, John, but we really rarely use interaction terms quite that way because it is difficult to achieve statistically significant interaction effects in a small population like this. So, we report that p and you are quite correct that the interaction term did not quite reach statistical significance but, nonetheless, the separation of the curves and the magnitude of the difference was quite striking. That is why it is hypothesis generating. Even if the interaction term had been significant, it would still only be hypothesis generating.

DR. TEERLINK: But I guess it would have lent more support to the direction of it in as much, as you are saying, it is difficult to achieve and it is difficult to achieve because of the large

range of variance within that estimate of the effects.

DR. COHN: Yes, I think that the interaction term would have been more important had we merely been looking at a wide range of samples in order to find some difference. This analysis was driven by pathophysiology, not by statistics. So, to find that there was a difference that came close to being statistically significant really confirmed the hypothesis rather than created one.

DR. NISSEN: Bob, you wanted to say something?

MR. SAMUELS: Yes, I had a question concerning accruals for clinical trials. I noticed the first two trials were conducted with all males in a VA environment, and I am curious about the third clinical trial in terms of female participation. How were those folks accrued to participate in the clinical trial?

DR. COHN: If you will wait for just a moment, I think Dr. Taylor is going to review all of that for you. I would rather have it held for

that.

DR. NISSEN: Dr. Temple, did you want to say something?

DR. TEMPLE: I just wanted to address the term "hypothesis generating" which is not quite precise. That is what we told people at the time when we said we are not ready to approve your application; those trials don't make the case but they are interesting and they do perhaps suggest a racial difference. But one implication certainly, and we conspired in this as you saw from the thing, is we thought an additional single trial that was persuasive--that is important, how persuasive it is--ould do the job. So, you are not starting from zero when you have a hypothesis formed; you are starting part way toward it.

I guess the other observation is that every time a statistician tells you that there was no interaction they always follow it with, "but, of course, the power of these tests is very low." But nobody said that today so I thought I would mention it.

DR. NISSEN: Just to comment on what Bob said, would we not view V-HeFT I and V-HeFT II very differently if the white/black curves were inverted? If we saw the opposite effect and it went in the other direction? I mean, yes, it is not the same kind of evidence as one would expect in a contemporary trial but it does go in the right direction. I think what you are saying, Bob, is the fact that it went in the right direction meant that you get some priors here before you actually do A-HeFT.

DR. TEMPLE: Well, some interesting questions have been raised about hospitalization, although I think after two years you are talking about very few people. In the first trial the white subset of the population looked like it was virtually nothing. I see there was 12 percent, or something like that but it is basically flat. The black patient was nominally significant. I know we don't take that p value seriously but, you know, that is not unimpressive. Then, in the comparison with enalapril in the white population it looks

like a placebo roughly, given the other trials, and the black population looks almost identical. So, how much to make of that exactly quantitatively is a very good question, but in a qualitative sense that didn't look too bad and that is what we thought.

DR. NISSEN: Milton, did you want to say something?

DR. PACKER: I wanted to make a follow-up point to John's question. It says the interaction value--well, not only is it a low powered test but the real question is what the sponsor always tries to do when they see things. They want to follow the signal. Following the signal doesn't depend on there being an interaction p value. What is really interesting here is that this is not a signal from one trial; this is a signal from two trials, two trials with very similar entry criteria. So, you have a better response to hydralazine and isosorbide dinitrate in V-HeFT I and V-HeFT II segregating according to the same baseline variable. So, you can't calculate an interaction p

value for that. You have consistency across two studies.

DR. NISSEN: I am going to move us on here unless the committee has further discussions. I think it was useful and important to get this all out on the table and we will have to all decide what weight we put on V-HeFT I and II in the discussions. And, Tom, you will have a chance to opine later.

DR. FLEMING: I will opine later. I think there is a lot more that needs to be said in terms of how you actually interpret interactions but, in essence, one has to be very cautious about a regression to the mean phenomenon, i.e., you specify to do (a); you then see in the data (b); and then you interpret (b) as though it was in fact a signal rather than noise in the data. In reality, it is probably both. So, in specific terms, if you see a global effect or lack of effect and that was your prespecified hypothesis, that is the most reliable interpretation.

Is it responsible to explore the data and

look for further insights? You bet. But as you are exploring, what you are finding is what is driven partly by signal but partly by noise. What draws your attention to it is that it looks different and in most cases when you see something like that it exaggerates what the true difference is. So, to back up what you are saying, you are probably exaggerating what the true level of interaction is when you are finding it through exploration in the data.

DR. NISSEN: I am looking forward to a lively discussion of these fine points of statistical analysis. We are now going to move forward--

DR. COHN: Let me just make one comment, Steve, because I want to remind you all that we are not focusing on the difference between blacks and whites. We are only focusing on whether this drug is efficacious in blacks. So, the interaction led us to exploring it but the question on the table is does this drug work in the population for which we have carried out A-HeFT.

DR. FLEMING: I was intending to stop but I can't stop--

DR. NISSEN: You can't stop, no!

[Laughter]

DR. FLEMING: Jay, you can't say "but we are intending to look in blacks." The blacks arose as an interaction term from the totality of the data.

DR. NISSEN: Let's move on. I can see we are getting into committee debate and discussion very early today, but let's hear all the data and I think we want to hear about A-HeFT and, Dr. Anne Taylor, I believe you are going to start.

DR. COHN: Right, and one personal comment before I sit down--

[Laughter]

DR. NISSEN: Gentlemen, this is not a debating society. We are trying to move forward!

DR. COHN: I want everybody to understand that I am due royalty payments based on sales of this drug, if it is approved by the FDA. I want you to understand that. I will now turn it over to

Anne Taylor.

A-HeFT: Rationale, Design and Demographics

DR. TAYLOR: Thank you very much, Dr. Cohn. Dr. Nissen, members of the panel, ladies and gentlemen, we will spend the remainder of the presentation discussing the primary registration trial, A-HeFT.

I would like to now direct your attention to a discussion of the design of the African American heart failure trial and to the characterization of the patient population in the study. Based upon the hypothesis generating subgroup analysis of the V-HeFT I and II trials, which suggested a particularly beneficial effect of combined isosorbide dinitrate/hydralazine in self-identified black patients with heart failure, the objective of the African American heart failure trial was to demonstrate the safety and efficacy of BiDil compared with placebo in black patients with moderate to severe heart failure who were concurrently receiving standard, guideline recommended contemporary heart failure therapy.

The study was designed as a double-blind, randomized, parallel-group, placebo-controlled trial in black patients with stable symptomatic heart failure while on standard therapy.

As mentioned earlier by Dr. Worcel, BiDil is administered as a fixed-dose combination tablet containing 20 mg of isosorbide dinitrate and 37.5 mg of hydralazine. Patients were randomized in a 1:1 fashion to receive BiDil of placebo added to standard, guideline recommended heart failure therapy.

In contrast to the V-HeFT trials, standard heart failure therapy in this trial included, but was not limited to, angiotensin converting enzyme inhibition, angiotensin receptor blockers, beta-blockers, aldosterone antagonists, digitalis and diuretics. All patients were started on one tablet three times daily and were force titrated to two tablets three times daily. The target dose was 120 mg of isosorbide dinitrate and 225 mg of hydralazine. These target dose levels were based on the actual doses achieved in the V-HeFT II trial

and the dosing interval was chosen based on the frequency with which patients actually took this combination in the V-HeFT trials. In addition, the target doses of isosorbide dinitrate and hydralazine were consistent with doses which have been demonstrated in the literature to reduce pre-load and after-load in the heart.

The study scheme of A-HeFT is shown here. A-HeFT was an 18-month study composed of a 2-week screening period, followed by a double-blind, randomized, parallel-group treatment period. Patients were required to be stable on standard heart failure medicines for 2 weeks prior to randomization. Stability was defined as no change in heart failure symptoms, signs or therapy, with no changes in weight greater than 2.5 percent. At the time of randomization a baseline echocardiogram and quality of life assessment were done, and basic chemistries and a hemogram were obtained. Randomization was stratified by beta-blocker usage and site by blocks of 8.

Telephone contact was maintained monthly

and study visits occurred at 3-month intervals, at which time quality of life assessments were obtained. At 6 months an additional echocardiogram was obtained. Patients were started on 1 tablet of BiDil 3 times daily, with a forced titration to 2 tablets 3 times daily initiated on days 3-5. Down-titration was permitted when patients could not tolerate the maximum target dose. Patients were followed up for clinical assessment of efficacy and safety every 3 months, as denoted in the study design slide.

All clinical events related to drug efficacy that occurred during the trial were reviewed by an independent central adjudication committee. Clinical events related to safety that occurred during the trial were periodically reviewed by the data safety monitoring board.

The key inclusion criteria for the trial are shown here. Patients were eligible for randomization if they self-identified as African American; had symptomatically stable New York Heart Association Class II-IV heart failure; and were on

standard background heart failure medications including neurohormonal blockade, digoxin and diuretics. If beta-blockers were included in a patient's background therapy, the patient was required to have been on beta-blockers for a minimum of 3 months. Left ventricular ejection fraction was required to be less than 35 percent or less than 45 percent with a resting left ventricular internal diameter of greater than 2.9 cm/m2, or greater than 6.5 cm absolute by echocardiography.

The key exclusion criteria for the trial are shown on the next two slides. The exclusion criteria applied in A-HeFT were comparable to those used in recent clinical trials in heart failure.

The primary endpoint for the A-HeFT trial was a novel endpoint used for the first time in this trial. It consisted of a score which weighted all-cause mortality, first heart failure hospitalization and change in quality of life at six months. Each element of the endpoint was assigned a score and the scores of the elements

were summed to yield a total score for each patient for the composite endpoint.

An important advantage to using this composite endpoint was that all patients contributed to the efficacy assessment irrespective of whether or not the patient sustained a clinical event.

Let's now look at the scoring system for the composite endpoint. This slide shows the primary endpoint scoring system used for the trial. Death at any time during the trial resulted in a score of -3, while hospitalization for heart failure received -1. Change in quality of life at 6 months was scored between +2 for an improvement of greater than 10 units and -2 for worsening in the quality of life greater than 10 units. Based upon this scoring system, the possible score for the composite endpoint ranged from -6, the worst score, to +2, the best score. If an element in the composite score was missing, the worst score for that element was assigned in computing the composite score for the patient. Therefore,

patients lost to follow-up would receive the worst score, a score of -6.

Quality of life was assessed at baseline and at 3-month intervals thereafter using the Minnesota Living with Heart Failure questionnaire. This is a 21-question patient self-assessment tool for measuring the emotional and physical effects of heart failure. Each question is scored from 0-5, with a total possible score ranging from 0-105. In evaluating the effective treatment on patient quality of life, please remember that lower scores indicate better quality of life.

The statistical plan for A-HeFT is shown on the following several slides. The primary endpoint analyses were performed in the intention-to-treat population which included all randomized patients. Once again, in assessing treatment efficacy, the worst possible score was assigned for any missing data, recognizing that the actual benefits of treatment might be underestimated when using this conservative approach.

As indicated previously, the categories involved in the primary endpoint were all-cause mortality, first hospitalization for heart failure and change in the quality of life at six months. Statistical assessment of the primary endpoint composite score was made according to the sequential methodology reported by Cui, Hung and Wangin 1999. The application of this methodology will be discussed further.

In addition to the primary composite endpoint, several secondary efficacy events were prospectively defined and assessed. First, all-cause death was assessed, including time to death as well as cause-specific mortality. Second, hospitalization for heart failure was monitored, including time to first hospitalization, number of hospitalizations and total days in hospital. Lastly, change from baseline in the quality of life was assessed throughout the 18-month trial duration at 3-month intervals.

The analyses for secondary endpoints were performed in patients who contributed data to the

study. Mortality and first hospitalization for heart failure were assessed using Kaplan-Meier log-rank testing. Between-group comparisons in mortality and hospitalization for heart failure were determined using 2-sample t-tests. Lastly, the change in baseline and quality of life was determined using 2-sample t-tests applied at 3-month intervals.

Because the primary composite score had not been previously used, the initial sample size of 600 patients was calculated by applying the scoring system to the V-HeFT database. There were a total of 2 prespecified interim analyses planned for safety and sample size estimation. The second interim analysis, performed according to the Cui, Hung and Wang method, was triggered at the juncture at which 50 percent, that is, 300 patients of the initial sample size had completed 6 months of treatment.

It should be noted, however, that at the time of this interim analysis 528 patients had actually been randomized. At that point in the

study it was determined that 900 patients were required for 80 percent power to achieve an alpha of 0.05 and that 1100 patients were required for 80 percent power to achieve an alpha of 0.02. The 0.05 significance was maintained but for robustness of potential results the final sample size was set at 1100.

This slide summarizes the events that led up to the early termination of A-HeFT. Mortality was not a prespecified primary endpoint in the trial and, consequently, no boundaries for stopping the trial based on mortality had been established before the initiation of the study. However, the data safety monitoring board made the observation that there appeared to be a treatment effect on mortality.

Based on this observation, boundary values for mortality were established using the O'Brien-Fleming type group sequential alpha spending function as described by Lan and DeMets. Treatment difference in mortality in March, 2004 fell below the value specified by these boundaries,

prompting the data safety monitoring board to recommend an additional safety review 3-5 months later.

Based on the positive mortality findings at this additional safety review, conducted in July of 2004, the committee made the unanimous recommendation to stop the trial. After favorable results had been shared with the steering committee and unanimous recommendations to stop the trial were received, NitroMed terminated the trial on July 19 of 2004.

I would like to turn your attention now to the details of the trial conduct and the key baseline characteristics of our patient cohort. A-HeFT was conducted in 180 sites, 169 of which randomized one or more patients. When the trial was stopped in July of 2004, 1050 patients were randomized to BiDil or placebo. No patient was lost to follow-up for vital status during the study. The first patient was enrolled in A-HeFT on May 29, 2001 and the study was terminated a little more than 3 years later, on July 19, 2004.

There were few differences between the two study groups. Notably, however, there were 40 percent women in this trial, the largest number of women in any single heart failure clinical trial. There were slightly more men than women in the placebo group. Importantly, there were no differences between the two study groups with respect to causes of heart failure. However, consistent with what has been observed in black patients in heart failure databases, ischemic heart disease accounted for only 23 percent of the heart failure in this trial. Hypertension and idiopathic causes accounted for the vast majority of heart failure in our study population.

As you can see from this table of baseline characteristics, the treatment groups were generally comparable with respect to hemodynamics and comorbidities. There was a clinically small but statistically significant difference in diastolic blood pressure between the groups, and a higher proportion of patients in the BiDil treatment group were diabetic.

This slide summarizes background medications of the study cohorts. There are two important findings to be observed. Firstly, there were no differences in cardiovascular medications between the two arms of the trial. Secondly, and most importantly, patients in this trial were well treated with neurohormonal antagonists. More than 90 percent were receiving ACE inhibitors or angiotensin receptor blockers; more than 80 percent were treated with beta-blockers; and approximately 40 percent were using aldosterone antagonists. A significant proportion of the patients were also taking digitalis and diuretics.

The patient disposition in A-HeFT is summarized here, and 1050 patients were randomized to BiDil or placebo. Of those randomized, a remarkably high proportion of the patients completed the study, 91 percent in the BiDil group and 86 percent in the placebo group. A relatively small proportion of patients were discontinued from the study and the majority of the discontinuations were attributable to patient death. Notably, the

vital status for all patients randomized in the study was known at study completion so that no worst case score was assigned due to loss to follow-up.

Study drug prescribed is summarized in this slide. Because the study was terminated early, patients randomized in the later stages of the trial did not complete the prespecified treatment duration of 6 months, and this is reflected in part in the drug exposure data. The mean number of days of treatment was 379 for the BiDil-treated group compared with 355 days for patients receiving placebo. The prescribed dosage of study medication was relatively consistent throughout the trial. The mean daily number of pills prescribed ranged from 4.4 to 4.9 BiDil tablets per day. Patients in the placebo group were prescribed a comparable number of tablets relative to the BiDil group.

The mean daily dose of isosorbide dinitrate ranged from 88-98 mg and for hydralazine from 188-199 mg, corresponding to about 4.4 to 4.9

BiDil tablets per day. These doses represent approximately 80 percent of the daily target doses for the two active components.

This concludes our discussion of the A-HeFT study design and population. Than you very much for your attention.

Questions from the Committee

DR. NISSEN: Let's take any questions for Dr. Taylor. Jonathan?

DR. SACKNER-BERNSTEIN: I would like to ask a question focused on the composite scoring system. I understand that it was applied retrospectively to the V-HeFT data to show that it did track consistently, but I am a little bit curious about how you determined the weighting of the different criteria.

My understanding would be that if you have a scale like this that is a composite where patient evaluation is an important part of it, the only way you can reliably interpret the information is to know that the weighting is consistent with a person's views. For example, how do you know that

a patient would say that a change in Minnesota heart failure score of 5-10 units is the same value as that patient suffering a first heart failure hospitalization that had met the criteria of an independent adjudication committee that confirmed that it was a heart failure hospitalization because you give them both a score of 1? If you are going to use the patient assessment for that Minnesota scoring system, then I think the value of the other components has to be validated, that the patient perceives them to be of equal value, otherwise you can't take the total number and start running it through a statistical test.

DR. TAYLOR: Yes, the scoring system was arbitrary. However, I think perhaps Dr. Ralph D'Agostino could talk about the application and the development of the scoring system and how it is applied to the V-HeFT database.

DR. D'AGOSTINO: Ralph D'Agostino, Jr., from Wake Forest University School of Medicine. In terms of the question you are asking specifically, I cannot probably help very much but the issue of

weighting each element of the score in terms of how it relates to the V-HeFT data--as Dr. Taylor has said, this is a novel composite and it was agreed upon as a primary endpoint prior to the initiation of the trial. The actual weights were mainly decided on through expert opinion among cardiologists as opposed to statisticians deciding upon the relative weight of each element.

As you can see though, each individual element was a secondary endpoint and analyzed separately. So, the issue of individual contributions was looked at in terms of how they contribute to the composite endpoint. In terms of validating it against V-HeFT data, as has been said, the data did not exactly match this new composite so we did our best, after the composite was derived, to see if it were applied to the V-HeFT data what kind of variability would exist in such a measure to help us understand the power and sample size needed.

DR. SACKNER-BERNSTEIN: So, when you analyzed statistically the result of the clinical

composite did you look at it as categorical outcomes or did you look at it as continuous variables?

DR. D'AGOSTINO: In this case we looked at it as a continuous measure, taking the range from -6 to 2.

DR. SACKNER-BERNSTEIN: I would think, and please correct me if I am wrong, that if you are using it as a continuous variable you have to know that the individual components are ones where each step of its component is one where that corresponds. Therefore, without validating it in a patient-centric way that a certain amount of change in the scale equals a certain amount of value to some of the other components, I would think that any attempt to look at this as a continuous variable would be statistically--I don't want to say invalid but questionable.

DR. TEERLINK: Just to add to that, as we noticed from the scoring system, if you had a 10-point drop in your Minnesota Heart Failure scale and were hospitalization, that is considered as bad

as if you died. Certainly, if somebody asked me whether I felt those were equivalent to a patient, you know, I wouldn't consider that equivalent.

DR. NISSEN: I never died so I really don't quite know how to assess that!

DR. COHN: Let me address that if I could because, as Anne has pointed out, this was arbitrary but it was based upon some data. We had actually gone back and studied a group of patients with severe heart failure--and this paper is published with Tom Rector as the first author, some years ago--in which we asked people to evaluate the relative importance of an improvement in quality of life by 10 units versus shortened life expectancy, which is as close as we could come.

The majority of patients said that a 10-unit improvement in quality of life was actually more important than life prolongation. In fact, they were willing to accept a shortened life for an improved quality. So, that helped us in scoring this. Yes, death is the worst and we didn't think that a 10-unit improvement was equal to death but

we did give it a score of 2. A 5-unit improvement has been documented to be associated with a general improvement in exercise performance and other positive features of life style. So, we thought that was worth 1. How you place hospitalization in that--remember that a hospitalization probably also will influence quality of life so they are somewhat interactive terms.

But in the long-run it is a very arbitrary scoring system and this was the first time it has been used. We felt it was as close as we could come to being rational.

DR. HIATT: I have read the Rector article a bit ago. Would you trade off that change in 10 for the rest of your life versus dying, as opposed to at one point in time asked on multiple occasions, having a blip in your worsening to then go to -10 and then with the possibility of getting better? So, I appreciate the attempt at the methodology but I think there are some concerns.

DR. COHN: Well, I would agree and this can be tweaked for future use. You will see the

data and I think it is all very consistent but there is no question this was an arbitrary attempt.

DR. NISSEN: We are going to discuss this further but, you know, one of the outcomes always of these panels is that everybody who is going to do trials in the future wants to watch, listen and sort of see what we can learn, and I think what you are exploring here is a new endpoint and I think we do have to explore it because it serves as a precedent for others.

I am going to jump in for a second and call on myself and say that I was a little bit surprised that you measured quality of life at a fixed point in time, namely at six months or earlier if that was the last measurement. I am thinking out loud, and I am not a heart failure doc but I kind of know that quality of life is something one experiences sort of every day. I was a little surprised when I read this that what you didn't do is repeatedly assess quality of life and then come up with some overall measure of what happened to quality of life during the course of

the trial. I would like to understand for my own benefit the rationale for picking a fixed point in time and saying how you felt at that point in time was the most important aspect in the trial.

DR. COHN: Well, this was a compromise, Steve, once again. We wanted to choose a time point where mortality would not impact upon the patient population. So, we chose a time point at six months which we felt was long enough to demonstrate the differences between the treatments but where mortality would not overwhelm the sample size so that almost everyone would be there for the six-month evaluation. So, it was an arbitrary decision.

DR. NISSEN: Tom?

DR. FLEMING: A question and a comment for Jay. The question first, so essentially the primary endpoint then is assessed at six months using quality of life status at six months, hospitalization and death status at six months?

DR. COHN: No, no--

DR. TAYLOR: Death throughout the trial.

DR. COHN: Morbidity and mortality was followed throughout the 18-month follow-up time.

DR. FLEMING: So, you did incorporate somebody who died at 12 months--

DR. COHN: That is right.

DR. FLEMING: --and was scored according to that. DR. COHN: They would get their death at 12 months and they would have their quality of life at six months.

DR. FLEMING: Just very quickly to add, because I think my colleagues have made some really key points here, I think we always run into complexities of interpretation when we have composite endpoints. A strength of this is that the components are each clinically relevant. A weakness is that they are clearly not of comparable clinical relevance. The words that I would use to express Jonathan's point are that if we are now using an analysis of variance and looking at the average increase, to what extent is there evidence, if what we have here is a 0.3 increase, that 0.3 increase is equally clinically relevant wherever it

occurs across the scale?

Just from a common sense aspect, I would also endorse what was said about how we look at the trade-offs. Clearly, I understand the argument that quality of life matters and somebody might well accept a risk or a trade-off of a somewhat shorter life for a much more substantial high quality of life while they are alive.

But just to create two scenarios, you could have a 10-unit improvement for a fairly short term and die and your score would be -1. It would be +2 on the improvement and -3 for death, versus somebody who has a 4-unit improvement and is hospitalized--that is one of the reasons it is 4 rather than 10--and lives are very long time. Those two people get the same score. I struggle with why that makes sense, that they are the same--somebody with a very short-term survival has a good improvement versus somebody else who also has an improvement and has a long-term survival would be the same score on this scale.

DR. NISSEN: That is why I was interested

in trying to understand the area under the curve. Again, quality life years is, in fact, a measure that we sometimes use in trials. I am sure you are more aware than I am of this. So, I am trying to get a sense of that because that is a very relevant consideration. In fact, it may be the most important consideration, that is, how much quality time can we give patients that have this very bad disease is very, very important.

So, I was going to give you a different scenario, which is somebody who is better at the time that this is actually measured but they are worse later, they are actually assigned a positive score. You know, you are assigned a score because you picked this one point in time and that is when you are going to assess it. It is pretty complicated to think this through. Bob, did you want to say something?

DR. COHN: Well, let me just point out, Tom, that the other side of the coin is just as difficult. If you are using mortality as a primary endpoint and you have a patient lying in bed on a

respirator, but who is alive, that patient doesn't count toward an endpoint and has the same score as a person who is healthy and robust at that time point. So, all of our endpoints are potentially flawed. We have put together here a collection of important clinical outcomes that we believe give you a little more insight into the overall effect of therapy. But it is flawed, of course, because of issues that you have raised.

Each of you has raised potential problems but you can do that with almost endpoint that you choose and I think there are always compromises in choosing an endpoint that you can use. Fortunately, we have elected also to look at each of these endpoints specifically, and you will see those data later. That, of course, is more powerful--the primary endpoint had to be established in order to achieve something that we said this is what we are aiming for; let's see if we can get there. But in terms of clinical insight, you will learn more from looking at the individual components rather than at the overall

score.

DR. NISSEN: I am going to let Dr. Temple jump in for just a second and then we will go to Bill Hiatt.

DR. TEMPLE: What is novel here is not interest in quality of life and importance but trying to put it on the same scale with other things, which people don't usually try to do. All these are imperfect and this discussion is very useful. But every composite endpoint has the same lesion. I mean, if you combine death with a new MI that isn't even painful, which is commonly done for all the 2b3a inhibitors, what is that? That is plainly 100 times worst than that but we count up all the total endpoints.

So, it is a very good discussion. This has been going on for a long time and we have long discussed, you know, giving 9 points for this one and 7 points for this one, and nobody is ever going to be entirely satisfied with it. So, I think the point that you can look at the separate endpoints is very helpful here. You know, composite

endpoints is the name of the game because you can't get enough deaths in trials, or you can't get enough of any one endpoint. So, we face this absolutely every time. Some attention was paid to trying to do it right. Whether you like the result is another question.

DR. NISSEN: I am going to go to Bill Hiatt next.

DR. HIATT: We are discussing the conduct of the trial now so I have several questions about the DSMB.

DR. PACKER: Could we just clarify one thing about the composite?

DR. NISSEN: Let's make it brief though.

DR. PACKER: The big concern I think the committee has is that there are lots of other pharmaceutical companies in the audience and they are looking at this primary endpoint and they are saying, "gee, you know, maybe we should do the same thing." I wasn't involved in the development of this primary endpoint. I will honestly tell you when I saw that I thought it was really weird. But

the reality is that the weights are arbitrary. There are all sorts of issues in terms of time--

DR. NISSEN: That is not a clarification point. You are kind of arguing the case here so--

DR. PACKER: Let me just say if the individual components didn't contribute, then there would be a real issue.

DR. NISSEN: Okay. We are really trying to get points of clarification and understanding rather than debate and I want to make sure we do that at this point in the morning. We will debate later.

DR. HIATT: So, the question is decision-making of the DSMB I think influenced the outcome of this study significantly. My first question is was there an original charter designed to just look at variance in the endpoint and then changing the sample size based on the overall variance, or effect size and variance? I have a series of questions. That is the first question.

DR. TAYLOR: Dr. David DeMets, who chaired the data safety monitoring board, is here with us

to discuss the proceedings of this committee.

DR. HIATT: Is it okay if we discuss the conduct now?

DR. TAYLOR: Yes.

DR. NISSEN: I think yes, that is what we want to do. We want to make sure we understand this and that is why I want the questions and the answers, not to focus on arguing the merits. Let's make sure we understand it all and then we will have plenty of chance to discuss what it means.

DR. HIATT: This is a key point of clarification so let's begin with the first question.

DR. DEMETS: I am Dave DeMets, University of Wisconsin. I had the challenge and the privilege of chairing and being the statistician of the data monitoring committee.

As you indicate, most monitoring committees are charged with monitoring trials for progress and safety and efficacy. This monitoring committee was given an additional charge to review the data because of the novel endpoint and using

the methods that Jim Hung and colleagues at the FDA developed to reassess and reevaluate the sample size. That procedure is based not only on the observed variability, but also the observed difference at that point in time and adjust the statistics and the penalties that go along with that.

DR. HIATT: So, you were looking at overall variance but also the effect size between group A and B.

DR. DEMETS: Yes.

DR. HIATT: Did you unblind yourselves to A and B or did you completely unblind at the beginning?

DR. DEMETS: At that point in time, which was March of 2003, we knew the data by A versus B. We did not know what A and B were, although it would be a stretch to say we didn't have a suspicion based on adverse effects, and so forth, but we did not formally unblind ourselves at that point.

DR. HIATT: When you began the DSMB, did

you start with a charter that was in place at the time that the DSMB was assembled?

DR. DEMETS: Yes.

DR. HIATT: And then somewhere in the middle after the second look you changed it. Correct?

DR. DEMETS: Changed the charter?

DR. HIATT: Did you change your charter?

DR. DEMETS: I don't know that we formally changed the charter; we introduced the issue of monitoring the mortalities more carefully.

DR. HIATT: Yes, so why did that happen? Originally the charge was to look at safety and look at sample size calculations. Of course, if the differences aren't going the way you thought and another few patients would change that, that would make sense to me. But then you actually changed it to open up the possibility of premature termination of the trial. It is just not clear to me why that happened.

DR. DEMETS: Well, any monitoring committee is monitoring for both safety and

efficacy. That is the challenge of all monitoring committees.

DR. HIATT: Well, no, I will ask another question. Was it really monitoring for efficacy? That is kind of a pointed distinction here. Most DSMBs monitor for safety and they often prespecify at the beginning that they are going to monitor for either futility or premature efficacy.

DR. DEMETS: There is a statement in the charter that talks about using O'Brien-Fleming boundaries to monitor the composite. They are pretty extreme boundaries as this procedure requires, and that was in place at the beginning for the composite. When we met in March of 2003 to do the sample size adjustment there were a few deaths--I think around 23 deaths, split 10 versus 13, as I recall--and we were again going to meet for our final interim analysis roughly a year later. So, at that point in time we had to discharge our sample size duties, if you will, so we met again in March of 2004.

When we met in 2004, as we were preparing

for that review, it was apparent that the mortality was picking up in terms of number of events, as well as the differential mortality between the two treatment groups. Because the protocol was silent about monitoring mortality--obviously, mortality was, you might say, the leading secondary endpoint--I felt as chair and as a statistician we had to do something at that point to guide us. I mean, I am one of those that believes that we don't have rules but we have guidelines and we always have to use our best judgment despite my passion for statistics.

So, we put in place the monitoring boundaries that were used for the primary composite, which is the O'Brien-Fleming boundary, using the methodology that I developed with a colleague, Gordon Lan, that allows you to be more flexible, using the same boundary type but using flexibility so that you could look when you needed to and as often as you needed to.

So, we took the boundaries that we had prespecified for the composite and applied them to

the mortality boundary, adjusting for the fact that we were now looking at a different point in time, and used those to help guide us. When we looked at the data at that March, 2004 meeting mortality was certainly trending but it didn't meet the criteria of that boundary although it was getting close. And we did all the other analyses--safety, adverse effects and so forth.

That was to be our last formal meeting, but we didn't feel very comfortable walking away from the trial at that point in time with this mortality trend, let's say, sitting there. The trial had nine months to a year to go before it would become completed and publicly available. So, we thought probably we should take a look again sometime in the middle. With summer vacations, and so forth, we finally settled on an early July meeting when we could get together by a conference call. So, that was the July 7th meeting of 2004 when we convened by a conference call. Dr. Ralph D'Agostino provided us with a report. That report had a mortality update. It had the adverse event

update. And we observed at that point in time that the deaths had now gotten even a little more significant and crossed the boundary which we had imposed at the prior meeting.

We did not at that particular conference call have all the analyses. We did not have the composite, for example, and we didn't have some other things. So, we sort of had to adapt quickly as to what to do so we put the meeting into recess, something that in 30 years of doing this I have never done before but you always learn in this business. We put the committee into recess for 48 hours until Dr. D'Agostino could have a chance to complete the analysis that we were looking for.

Yes, the mortality crossed the boundary but, for the reasons you suggested, it wasn't prespecified. In some sense, it wasn't good enough just to look at mortality so we asked how does the rest of the data look. So, that was what we were provided 48 hours later. We looked at the composite for all the pluses and so forth that you talked about. It was significant. Time to first

hospitalization was significant, as well as worsening of heart failure. Then we looked at the conventional subgroups. Most of this, of course, is coming up in Dr. Yancy's presentation so I won't go through all that. But we looked at the package. The primary was significant even at the O'Brien-Fleming boundaries. The subgroups were consistent. The other outcomes were consistent and mortality had crossed that boundary.

So, at that point in time, on July 9th I guess of 2004, we felt that with everything that we had seen it was necessary to recommend to the steering committee that we should terminate the trial. We were convinced it was the mortality effect and the composite effect, and a consistent effect. I may have more than answered your question--

DR. HIATT: No, I think the process of your thinking is absolutely critical to understand, and it sounds like you were kind of adapting as you went along. It sounds like the mortality signal looked like a safety signal to you.

DR. DEMETS: You know, mortality and safety are two sides of the same coin. It is safety for one, efficacy on the other side. So, it is probably the only outcome you can say that about. Right, most monitoring committees have certain guidelines and what you learn is that even on the best day they are not always adequate for everything you are going to face. So, you have to be a bit flexible, always keeping in mind the best statistical principles, clinical trial principles, common sense, good physiology, good medicine, and all that stuff.

DR. NISSEN: I do agree with the statement that when common sense tells you--even though mortality is not the primary endpoint here, when you see something very strong on mortality there is obviously an ethical and a moral responsibility to make a decision on that, and I think that the committee extended its charter because of that ethical responsibility and I completely understand your thinking about that, and that makes some sense. Dr. Ota Wang?

DR. OTA WANG: I actually want to shift gears. I have questions about your inclusion criteria. It was my understanding from previous statements that the underlying assumption of this study is really looking at biological differences between your black and your white sample group and that is why you actually developed the A-HeFT study.

So, I have a two-part question. One, I would actually like to hear more and have you discuss your scientific rationale on evidence about how your self-reported racial identity is related to this biological assumption. Two, I am actually wanting to hear more reasoning for using a self-reported racial category rather than a more direct biological, quantitative trait as an inclusion criterion.

DR. TAYLOR: I will answer and then I will turn the microphone over to Dr. Sabolinski. Our hypothesis was not the differences between the two races but was to confirm the hypothesis suggested in V-HeFT that there was particular efficacy in

this group that self-identifies as African American. We did not have any other biologic markers, other than self-identification and the evidence in the literature which suggests that individuals who self-identify manifest differences in the presence of hypertension, the target organ damage and in the causes and outcomes of heart failure.

DR. OTA WANT: So, if I presented myself and said I am black, would you allow me to participate in the study?

DR. TAYLOR: Yes. Self-identification was the criterion.

DR. OTA WANG: So, there is a possibility that there are people who look like me or other people around the table who were included in the study based on their assumption of who they are and not what you are presuming they should be.

DR. TAYLOR: Yes--

DR. OTA WANT: Okay.

DR. TAYLOR: --we allowed only self-identification as the criterion.

DR. SABOLINSKI: I would just like to make an additional comment. That is consistent with the census that black and African American are both considered synonymous. Also, it is consistent with FDA guidelines with regard to collecting ethnicity or race in clinical trials, that black or African, self-identified African Americans is the method of doing this.

DR. OTA WANG: Absolutely. I guess since I am at the Genome Institute as well, when we start talking about racial classification a lot of it is just based on what people look like and assumptions people make. For racial categories I think people use skin tone and the racial categories are a proxy for skin tone, and I don't think skin tone is necessarily a great proxy for a biological sort of trait. What we do know is that there is a lot of within-group variation between groups. So, I guess I am a little wary because I am not really sure who you mean by black.

DR. NISSEN: I was actually going to follow on with that just a little bit and maybe get

a clarification. I asked Jay earlier about this question about what is it that is different. I guess what I would like to understand is whether it is possible to test for this deficiency in nitric oxide production that has been talked about. Obviously, if you could actually have a direct marker of who would benefit maybe there are some Caucasian or white Americans that would have this trait and would fall into the same group. Is that possible?

DR. SABOLINSKI: The company is working on direct assays for nitric oxide. There are no predictors right now or assays available and we are committed to expanding the population by looking at various physiological, functional and genomic markers that would expand the population in the future.

DR. OTA WANG: I guess my point is if we start looking at the literature on evolutionary and genetic history, depending on how far you go back, we will all eventually originate on the Continent of Africa. So, when we start talking about genetic

markers I think we need to start thinking about what do you actually mean about how it is related to identifiers that I think are much more socially and politically sort of designated. So, I think we need to keep that in mind because I think there is a presumption here that somehow this self-identified social identifier is somewhat equivalent or representative of a biological process, and I am not sure it really is.

DR. NISSEN: Jonathan?

DR. SACKNER-BERNSTEIN: As far as the study design criteria that led to you selecting the African Americans as a population of interest, I think it would be potentially enlightening to explore one of the other subgroups that came out--I have seen this at least in V-HeFT II and maybe it holds in V-HeFT I as well--of another group that is likely to have nitric oxide deficiency and that is the subjects who were alcohol users, abusers, heavy users--it has been labeled different ways in documents. Alcohol abuse has been reported to deplete nitric oxide. There was a higher

percentage numerically in V-HeFT that were heavy alcohol users than there were African Americans. I think it was about 33 percent versus 26 percent, something like that.

So, I wonder whether we are actually finding out that this is a drug biologically that is useful in patients who have a nitric oxide deficiency but perhaps the real hypothesis is related to alcohol use. I am wondering whether that was information that was collected in A-HeFT.

DR. SABOLINSKI: The signal that we followed was followed from V-HeFT I to V-HeFT II, as Dr. Cohn has explained, and was based on survival data. We did do subset analyses and those analysis showed that race was the strongest predictor of survival. Dr. Cohn has shown this in V-HeFT I, which was placebo-controlled, and in V-HeFT II. In A-HeFT we did not do a subset analysis to generate a point estimate for alcohol use.

DR. SACKNER-BERNSTEIN: The V-HeFT II data actually show the same kind of difference in

responses by alcohol use as they do by race. So, in patients in V-HeFT II who were alcohol users--I don't remember the exact term--there was equivalent effect, whereas in patients who were not alcohol users the outcomes were better on enalapril. The point estimates are roughly the same.

DR. NISSEN: Is that published stuff?

DR. SACKNER-BERNSTEIN: That is a question I asked to the FDA when I was reviewing stuff. That was in the FDA review document for '97.

DR. PACKER: Was that based on hazard ratios or a comparison of annual mortality rates? Because prior to the recent submission all subgroup analyses were based on the paper Jay authored that used annual mortality rates and not hazard ratios.

DR. SACKNER-BERNSTEIN: I will tell you that in a second--these are crude mortality rates based on race and alcohol use.

DR. PACKER: Right, you have to be very careful here. It would be much better to look at this based on hazard ratios. They don't come out the same way if you do it using hazard ratios.

DR. SACKNER-BERNSTEIN: Well, I have the 95 percent confidence intervals on hazard ratios too and those look roughly the same.

DR. NISSEN: That is a very interesting point. We will come back to that. We have no further information here. One more question.

DR. KASKEL: Right, 16 percent of the study group and 18 percent of the placebo group had renal insufficiency. Did you break this down into the subgroups of the Doci criteria for renal insufficiency, and do you have data on baseline protein and microalbumin excretion? Nitric oxide has been shown to change in patients with chronic kidney failure in their kidneys. This is important.

DR. TAYLOR: No, we had only baseline creatinines. We did not have measurements of albumin.

DR. KASKEL: Were the creatinines broken down for the different groups for the Doci guidelines of chronic renal insufficiency?

DR. TAYLOR: No, we used separation of 2

to indicate renal insufficiency but did not break down.

DR. NISSEN: If there are no further questions I am going to ask the sponsor to move on with the next presentation, which I believe will be Dr. Clyde Yancy on results and conclusions.

A-HeFT: Outcomes

DR. YANCY: Good morning, Dr. Nissen and good morning to the panel members, and thank you, Dr. Taylor.

I would like to present the clinical efficacy and safety findings of A-HeFT. Let's begin with an examination of the efficacy of BiDil. As you have already heard, the primary endpoint was a composite score derived from mortality, hospitalization and quality of life data. For this primary composite endpoint there was a significant difference in favor of BiDil, -0.16 in the BiDil group versus -0.47 in the placebo group, with a p value of 0.016. The difference in favor of BiDil was apparent for each component of the composite score and, thus, all three components contributed

to the success of the primary endpoint.

As you know, each component of the primary composite endpoint was specified individually as a secondary endpoint. Let us look at each component, first at all-cause mortality. Overall, during the course of the trial 54 patients, or 10.2 percent, in the placebo group died compared with only 32 patient deaths, or 6.2 percent, in the BiDil group. Time-to-event analysis revealed a 43 percent lower risk of death in the BiDil group when compared with the placebo group, p equals 0.012. The reduction in risk was primarily related to reduction in mortality due to heart failure deaths.

When we look at specific causes of death, the total number of heart failure deaths was significantly lower in the BiDil group than in the placebo group, p equals 0.012. There was also a significantly higher risk for pump failure death in patients treated with placebo plus standard heart failure therapy. No other notable differences in cause-specific mortality were observed between the groups.

Let us look at hospitalization for heart failure. Overall, during the course of the trial 130 patients, or 24.4 percent, in the placebo group but only 85 patients, or 16.4 percent, in the BiDil group were hospitalized for heart failure. Time-to-event analysis revealed a 39 percent lower risk of hospitalization for heart failure in the BiDil group, p less than 0.001.

It is important to clarify the hospitalization data. The event rate for first hospitalization for heart failure is significantly lower in the BiDil group versus the placebo group. For those patients who were hospitalized, the mean number of total hospital days for the duration of the trial is not different. These data were previously communicated to the agency and have been presumed to mean that the difference in total hospital days was insignificant. However, these additional analyses of heart failure hospitalizations, including protocol specified secondary analyses, demonstrate the total impact of BiDil on hospitalization. Compared with

placebo-treated patients, BiDil-treated patients experienced a significant reduction in the number and duration of hospitalizations for heart failure.

Because death and hospitalization represent competing risks, we carried out an analysis of the combined risk of death or hospitalization for heart failure. BiDil reduced this risk by 37 percent, p less than 0.001.

Finally, let us look at quality of life. Please remember that a reduction in score means an improved quality of life. At 6 months, the time point prespecified for the primary analysis, there was a 7.1 point improvement in the BiDil group compared 3.1 point improvement in the placebo group. The p value for this difference was 0.011. The benefits of BiDil on patients' quality of life were seen not only at 6 months but also at earlier visits, at later visits and at the last double-blind observation.

Subgroup analyses for the primary composite endpoint revealed consistency of the treatment difference across subgroups not only for

demographic and clinical characteristics, but also for background medication use. The subgroups that did not show a favorable point estimate were those that were very small.

The same was true for all-cause mortality. Again, there was consistency across the subgroups not only for demographic and clinical characteristics, but also for background medication use. The subgroups that did not show a favorable point estimate, again, were those that were very small.

Finally, there was consistency of the treatment difference for hospitalization for heart failure for subgroups defined by clinical characteristics and also for background medication use. The subgroup that did not show a favorable point estimate was among the smallest.

This slide shows the changes in blood pressure in the placebo and BiDil groups. Blood pressure increased by about 1 mm Hg in the placebo group but decreased by approximately 2 mm Hg in the BiDil group. However, the mortality effects of

BiDil did not appear to be related to its blood pressure lowering effects. When the mortality estimates were adjusted for baseline blood pressure and change in blood pressure, there was little shift in the magnitude of the estimated treatment effect.

This slide summarizes the main findings of the A-HeFT trial. In this trial BiDil decreased the risk of mortality by 43 percent; decreased the risk of first hospitalization for heart failure by 39 percent; and improved patient quality of life over the duration of the trial. The benefits were preserved or maintained in patients receiving various iterations of standard heart failure therapy and in patients with different comorbid conditions.

Let me now summarize the safety data in the A-HeFT trial. On this slide we have listed the common adverse events observed in A-HeFT, headache, dizziness hypertension were more common among BiDil-treated patients, whereas worsening heart failure was more common in placebo-treated

patients. When we looked more closely at headache, dizziness and hypertension, these events did not generally lead to the discontinuation of treatment, although discontinuations for these three reasons were more common in the BiDil group.

This slide tabulates the most common serious events. The only event to show a between-group difference was worsening heart failure which occurred more frequently in the placebo group.

The discussion regarding adverse events is important as the patients experiencing adverse event contribute substantially to the cohort excluded from the per-protocol analysis. This limited group of protocol specified and strictly defined patients constituted only 40 percent of the patients studied. The findings in this group on the composite endpoint were not significant but, importantly, the number of events in this group was too low to carry meaningful statistical validity.

In conclusion, this slide reminds us again of the main findings of the A-HeFT trial. In this

trial BiDil decreased the risk of mortality by 43 percent; decreased the risk of first hospitalization for heart failure by 39 percent; and improved patient quality of life over the duration of the trial. BiDil was generally well tolerated in the proposed population and for proposed use.

Let me remind the members of the panel that the findings of A-HeFT are important. Heart failure is a pressing cardiovascular illness and BiDil represents a new treatment for heart failure as it affects African Americans. African Americans experience heart failure at a greater frequency and have an unusual natural history. The disease occurs earlier, oftentimes with more advanced left ventricular dysfunction and with a definite variance in the presume etiology of left ventricular dysfunction. Of great concern is that clinical outcomes are more troublesome and the prognosis is less favorable. BiDil, when given to African Americans with heart failure, improves outcomes. In my judgment, this is a benefit that

we must extend and an opportunity that we shouldn't miss.

Thank you, ladies and gentlemen of the panel. I would now like to turn it over to Dr. Packer to complete our presentation and to accept your questions.

DR. NISSEN: We are actually due for a break and I would like to actually take it. So, no more than 15 minutes. Fifteen minutes from now we are going to start with committee questions for Dr. Yancy.

[Brief recess]

Questions from the Committee

DR. NISSEN: Let's come to order. All the side bar conversations need to cease and we need to all sit down and get to work. Dr. Fleming and Dr. Temple, please take your seats. I am going to get started. Robert Samuels, our patient representative has a question for Dr. Yancy. Please proceed.

MR. SAMUELS: Thank you. Dr. Yancy, as a potential user of this new product I am very

curious about how do you measure quality of life as you put forth in your data?

DR. YANCY: Thank you very much for the question. We use the Minnesota Living with Heart Failure questionnaire, which is a 25-question survey that receives a numerical score from 0-125. In our study the responses to that questionnaire gave most patients a mean score of 50, which is consistent with pretty symptomatic heart failure. That was the baseline and we measured the change from that baseline, and we did so at baseline and at three-month intervals during the trial, with the six-month measurement being our prespecified measurement that was built into our endpoint.

DR. NISSEN: Other questions from the committee related primarily do Dr. Yancy's presentation? Dr. Sackner-Bernstein?

DR. SACKNER-BERNSTEIN: I have a couple of questions. One relates to some of the data about heart failure hospitalizations because I may just be misunderstanding some stuff, but it looks like the data in your slides might be different than the

data in the review. In the slides you report things--and I can tell you, it is slide CE-12 that I am talking about where you talk about heart failure hospitalizations--

DR. YANCY: Sure. Can we put that back up, please?

DR. SACKNER-BERNSTEIN: --and you show some very important things favorably affected by the combination of hydralazine and nitrates, including things such as, near the bottom, the mean number of days in hospital for heart failure per patient. But if I look at the FDA briefing document on page 20, I get a very different impression because here days in hospital for heart failure hospitalization per patient has a mean of 13.7 for hydralazine nitrates compared to 15.3 for placebo, with a p value of 0.54. Could you help me understand the differences and what I am misinterpreting here?

DR. YANCY: Absolutely, Dr. Sackner-Bernstein. The data that you are referring to in the FDA briefing document reflects a

conditional analysis, the condition being having been hospitalized. The event of hospitalization occurred in 130 patients in the placebo group and 80 or so in the BiDil group. So, that number that you see there reflects two things, one, the duration of hospitalization for those patients who were actually hospitalized, but it is also a mean of the aggregate number of days of hospitalization for those patients during the duration of this trial.

DR. SACKNER-BERNSTEIN: That is great; that is very helpful. How about data on cardiovascular hospitalizations in A-HeFT? Can you share those data with us?

DR. YANCY: let me refer that question to Dr. Chris O'Connor who is here from our adjudicative events committee.

DR. O'CONNOR: Chris O'Connor, from the events committee. What is the specific question?

DR. SACKNER-BERNSTEIN: Well, the first one was total number of cardiovascular hospitalizations between the groups, not just heart

failure hospitalizations.

DR. O'CONNOR: Do you have that data on a slide?

DR. YANCY: While they are getting the data up, let me just remind you that the analysis of heart failure hospitalizations plus other cardiac hospitalizations was less significant, as you have probably already seen. It was the heart failure hospitalizations that was more significant.

DR. SACKNER-BERNSTEIN: And I think that would be reasonable. I just want to make sure that it is not data going in the wrong direction as opposed to diluted.

DR. O'CONNOR: I mean, if you look numerically on page 28--

DR. NISSEN: Do you have a slide? If everyone in the audience could also see would be great.

DR. SACKNER-BERNSTEIN: I would like to just see the results the way you presented it, according to CE-12 at least, and have some general idea of that per patient. Because this, on 20, is

for the people hospitalized and there is a differential risk over time with some things like that. So, on a per patient basis what were the cardiovascular hospitalizations between groups?

DR. O'CONNOR: We did not prepare an analysis for cardiovascular hospitalizations in the same way as you have seen for heart failure hospitalizations.

DR. SACKNER-BERNSTEIN: How about total number of days in the hospital between groups? Again, I wouldn't necessarily expect there to be a difference but I would like to make sure that it is not going in the wrong direction.

DR. YANCY: Well, what is interesting is that if you look at the analysis of total days of hospitalization that actually is statistically in favor of BiDil treatment.

DR. FLEMING: Just on this point, it would be very helpful to see--if you don't have it now, to get it over the lunch break, to be able to show us a slide that gives time to hospitalization, all-cause hospitalization and total days of

all-cause hospitalization.

Just to follow-up on Jonathan's point, it is not that we would expect statistical significance there, but I would just want to make sure that the magnitude of the difference isn't diluted. If we went back, for example, to last year's oncology drugs advisory committee for review of Olympta[?] in lung cancer, the sponsor focused on febrile neutropenia-related hospitalization and showed a big difference, but when you looked at all hospitalizations it was the same. So, I just want to be reassured that if we are going to infer from these data a quality of life benefit through a reduction in the hospital, as you are correctly noting, the signal for that is probably most sensitively addressed by looking at heart failure hospitalization days. And I just want to make sure that total hospitalization days doesn't neutralize that.

DR. NISSEN: Actually, it is interesting you should suggest that because, I mean, every now and then we see that there are competing risks

where a drug will reduce hospitalization for one cause and increase it for another, and we need to know that. We have actually seen this several times at this very committee. So, you know, it just makes us more comfortable when we have all the data and we can look at it very carefully and convince ourselves that we are not missing something, and it is our job to probe you. Jonathan?

DR. SACKNER-BERNSTEIN: On slide CE-22 where you show changes in blood pressure over time--

DR. YANCY: Can we put CE-22 up again, please?

DR. SACKNER-BERNSTEIN: I think we are given the impression that the blood pressure goes up with hydralazine relative to BiDil by this graph. But that was certainly not your intention in showing the slide. Right? Because blood pressure actually is lower over time. Those p values you are reporting are just compared to baseline. Right?

DR. YANCY: They are compared to baseline.

DR. TEERLINK: I was noticing--

DR. ARCHAMBAULT: Tad Archambault, from Virtu Stat. A point of clarification, I think those p values are for between treatment comparisons at each time point and what is shown on this slide is the change at each time point.

DR. NISSEN: Okay. Dr. Teerlink:

DR. TEERLINK: I am looking at page five of the statistical review, and the comment there is that there was no statistically significant finding on any of the other five secondary endpoints, which I presume include the echocardiographic, structural remodeling endpoints and the BNP endpoints. But I don't see any of the data for that anywhere. Are we to presume that those were negative? The statistical reviewer said there were no changes in those.

DR. SABOLINSKI: There was agreement with FDA that ejection fraction and BNP would not be part of this file and would be submitted at a later date.

DR. NISSEN: Tom?

DR. FLEMING: But does that mean that we shouldn't at least discuss the nature of the results if there is no data available on those measures?

DR. NISSEN: I think we can certainly ask the sponsor to see the data. It may not be part of this official filing but it is certainly part of the study as prespecified I think.

DR. TEERLINK: Dr. Cohn has always been a big proponent of kind of putting together the pathophysiology changes--

DR. NISSEN: No, I mean is it okay if we ask for that? There were some secondary endpoints that were in the original design that we don't see here and several people are looking to see that. Even though we understand that you agreed that it wouldn't be part of this official filing, if the data are available I think it is reasonable for us to see them.

DR. SABOLINSKI: If I may, the information or the conclusion that you referred to is in Dr.

Hung's review. We have not completed those analyses and they are not available for ejection fraction and BNP.

DR. NISSEN: You know, I understand, although it is a little puzzling in that the trial was completed sometime ago and BNP is a laboratory measure that we actually get in five minutes at the bedside in our CCU and even in the laboratory they get it back in an hour. So, it is not difficult and it makes us uncomfortable when we have prespecified endpoints and we don't get to see the data. Tom, I don't know how you feel about that.

DR. FLEMING: Yes, I am very pleased that the sponsor has put forward what I would consider to be the most clinically relevant measures here, the measures on mortality and hospitalization and even quality of life. These other measures are predominantly biomarkers. Nevertheless, having said that, I would generally expect much greater sensitivity on the biomarkers than on the clinical endpoints and when the statistical review says these weren't significant it just sends up a bit of

red flag to me as to what is happening with those biomarkers?

DR. SABOLINSKI: Once again, those data were not presented to the statistical reviewer. If I may, our analyses are not complete. We have looked at the overall trends. We do see an improvement in ejection fraction directionally in favor of BiDil at six months, and we do see a BNP decrease in BiDil-treated patients when compared to placebo at six months. We do plan on submitting these data to FDA when our analyses are completed.

DR. NISSEN: The other reason we like to look at these things is that some of us actually think--Tom doesn't but some of us think biomarkers are occasionally useful. When you do a study like this, which is really a kind of a landmark study, it helps us to validate to understand, you know, whether biomarkers are sending us consistent signals that we can look at in other pilot studies in other hypothesis generating ways. So, it is just useful for the scientific community when we have those analyses. Norman?

DR. STOCKBRIDGE: Just a suggestion, the other reason why you might care a little about this is that you may want to pick one of the secondaries to pay specific attention to, and understanding how what the rules were for interpreting the secondary endpoints may be important to you.

DR. NISSEN: Well, i was getting to that actually because there is a multiplicity issue here. I knew Tom would raise it even if I didn't. You know, I have a view to how one reports clinical trials, and one of the views that I have is that if you predefine an endpoint as primary, secondary or tertiary, then when you report the trial you report all the endpoints, the good with the bad and the indifferent, and you get a more complete picture of what actually happened. So, that level of transparency always makes me more comfortable and then I can do whatever statistical corrections, or ask my friend, Tom Fleming who can do more on the back of an envelope than I can with a computer, to actually compute all of this. And we don't get that opportunity here and that is a problem.

DR. TEERLINK: I think the other reason I was bringing it up is that we are being asked to basically interpret this on the basis of one trial. So, the kind of consistency of the evidence, the direction of all these things would have been useful and supportive. If, in fact, they weren't I am now kind of left--I am not sure how to interpret them.

DR. ARCHAMBAULT: A point of clarification, in the statistical review, on page five, the sentence that was read that there were no statistically significant findings in any of the other five secondary endpoints--I think we took that as Dr. Hung referring to his review and additional analyses that we had performed and not necessarily identifying the secondary analyses that had been pointed out in the protocol. Those secondary analyses were performed, many of which were considered sensitivity analyses, and those in fact were statistically significant in favor of BiDil. We don't want people to get the impression that all secondary analyses failed statistical

significance.

DR. NISSEN: No, I understand but we are not seeing it all and, you know, part of what goes on at these panel meetings is that other people who are developing drugs want to listen to what went on here because it does tell you what a committee like this likes to see and why it is that we want to see as complete a picture of the results of a trial as we can. I think the discomfort of the committee should be noted. It doesn't necessarily mean that it is an approvability issue, but it is a confidence builder when you get to see everything and I think that this should be duly noted by everybody. Other questions?

DR. SABOLINSKI: Well, if I may go back to the question that you asked about hospitalization and directionality, we do have a slide for that and I would like to call it up.

I am sorry that it is in this format but it is taken from our clinical study report and it is not in either briefing document. What it shows is that for hospitalizations for all-cause you see

a p value of 0.4. It is in the direction of favoring BiDil.

For hospitalizations for other cardiac causes we do see a p value of 0.5. It too is in the direction of favoring BiDil. For hospitalizations for other or non-cardiac causes we see a p of 0.7 and nominally the numbers do favor BiDil. Overall, there are no statistically significant differences but this does show the trends which you asked about.

DR. FLEMING: Could you clarify, just so we can put this in context with what you showed earlier? Earlier you gave us data on heart failure hospitalizations and you were addressing, I think, in that context an issue raised by the FDA review when they said the total days in the hospital weren't different. If I followed, you gave an appropriate response to say the average days in the hospital when you had a hospitalization weren't different but the total days were different because you had fewer hospitalizations--

DR. SABOLINSKI: That is correct.

DR. FLEMING: --and you gave a statistic that was 173 versus 251?

DR. SABOLINSKI: Yes.

DR. FLEMING: How does that relate to the 202 and 221? Because, in fact, those numbers, when you go to all hospitalizations, you would expect to be larger than the 173 and 251 numbers.

DR. SABOLINSKI: These are patients.

DR. FLEMING: The other numbers weren't patients? So, it was just total numbers of heart failure hospitalizations.

DR. SABOLINSKI: That is what the other analysis showed, and in the middle portion of the slide we do show days in hospital, days per patient--

DR. FLEMING: For the moment I just want to kind of drill down on one issue at a time. So, the numbers that you gave us before were then total number of heart failure hospitalizations, not total number of people with a heart failure hospitalization?

DR. SABOLINSKI: We have actually shown

things two ways. One way was by showing the patients who had a heart failure hospitalization and all the days that they were hospitalized--

DR. FLEMING: No, I understand--

DR. SABOLINSKI: --and showing a mean.

DR. FLEMING: So I can put this number into context, I am just looking for if you gave us the total number of people with heart failure hospitalization because the 173, 251 you are now saying are the total numbers of heart failure hospitalizations. What is the number of people by treatment arm with heart failure hospitalization?

DR. ARCHAMBAULT: It is 85 and 130, 85 for BiDil and 130.

DR. FLEMING: So, 85 and 130. So, as we look at that data--I am heading somewhere--down the road, to me, this is going to be a critical aspect in my own judgment about strength of evidence. How strong is the evidence beyond mortality specifically on quality of life interpreted through hospitalization? And those data are, in my view, stronger than the mortality data, and it is 85

against 130 so it is a differential of 45 hospitalizations for heart failure and, yet, the critical number you are telling us is that the 85 grows to 202 when you look at all-cause and the 130 grows to 221. So, this excess of 45 is cut in half when you look at all people who have hospitalizations.

DR. ARCHAMBAULT: May I point out though that on the first one where we did heart failure hospitalizations, the 85 versus the 130 was statistically significant. The conditional analysis for those patients who were hospitalized in terms of mean number of days was not significant. Here you have pointed out that the increase from 85 to 202 seems large relative to the 135 to 221. However, the conditional analysis here on the average number of days in hospital for these patients is now statistically significant which counterbalances that to a certain point.

DR. FLEMING: Basically, I am just simply trying to look at your primary endpoint, time to first heart failure hospitalization, and the

corresponding number of people with a heart failure hospitalization at 85 and 130 which, to my way of thinking, is the strongest signal for efficacy that you have in this trial. But then I am trying to say in the context of all-cause--I don't care if it statistically significant; I just want to know that that excess is maintained. It appears that half of the excess is maintained. It appears that way.

DR. ARCHAMBAULT: It is 202 to 221. The mean days of hospitalization is 13 compared to 17.7 and that is statistically significant at 0.12.

DR. FLEMING: But an order of magnitude less significant than your analysis of time to heart failure hospitalization.

DR. ARCHAMBAULT: That is correct.

DR. PACKER: Tom, you also have the issue of competing risk of mortality.

DR. FLEMING: Sure, I understand.

DR. NISSEN: Other committee questions? I am sorry, Tom. Please go ahead.

DR. FLEMING: Just one other feature of this, as I am trying to get a more global sense of

the impact beyond mortality, unscheduled ER visits--can you show us those results?

DR. YANCY: Are we able to pull up the unscheduled ER visit data? While that is coming up, I think you have data in front of you that demonstrates that that was not significantly different.

DR. FLEMING: Pardon?

DR. YANCY: I think you have data in front of you that demonstrates that the unscheduled ER visits were not statistically significant.

DR. FLEMING: So, while you are pulling it up maybe I can go to another question and we can come back to it after you pull it up. As I understood the analyses--just to come back to Steve's point--I interpret p values--in fact, there is only one p value I understand--only one, and that is the p value that corresponds to the prespecified primary analysis of the prespecified primary endpoint. If that is 0.25 less, then what I am saying is that that is something sufficiently unlikely by chance alone if there was no effect

where I can infer causal benefit or treatment-induced effects. Everything else is not irrelevant but much more difficult to interpret.

Am I correct to understand that the actual prespecified primary analysis was a per-protocol analysis, or was it in fact the ITT analysis? I am asking because it was raised in the FDA review. Could you clarify this per-protocol versus ITT?

DR. SABOLINSKI: Yes, in conjunction with FDA we did agree on the composite score and that composite was the primary efficacy analysis in the trial and the population is the intent-to-treat population. So, that was the prespecified primary efficacy endpoint and the primary composite score, made up of three components, was the primary endpoint in this study for the intent-to-treat population, all 1050 patients.

DR. FLEMING: So, looking forward to a question that we are going to be asked to answer so that you can help with the insight for that, it says, however, the sponsor's prespecified per-protocol analysis is not significant, p of

0.46. Why were 60 percent of people excluded from the prespecified per-protocol analysis? As we prepare to answer that question, can you provide us clarity as to your insights about this?

DR. SABOLINSKI: Well, first I think that had we shown results that were statistically significant in the per-protocol population and not in the intent-to-treat population we probably wouldn't be having this conversation.

DR. FLEMING: You are absolutely right.

[Laughter]

--While I completely agree, I just want to understand the context of what you planned and what you saw. So, so we can answer this FDA question, can you clarify exactly what you had said you would do with the per-protocol and then what actually you saw with the per-protocol?

DR. SABOLINSKI: I believe that Dr. Yancy did show the slide that referred to the number of events, deaths and heart failure hospitalizations, that were eliminated. There were approximately 300. Given the small number of patients that were

left in per-protocol and the small number of events, the p was not significant. However, it was in the same direction.

DR. FLEMING: So the per-protocol population does, in fact, delete--we see 41.7 and 40 so it does delete about 60 percent of the people.

DR. SABOLINSKI: Yes.

DR. FLEMING: Two questions, exactly what did your protocol state the role of this per-protocol analysis to be and, secondly, can you show us at least what the results looked like for that analysis? I know the p is 0.46 but when you say it wasn't significant, lack of significance can occur because you see a nice difference but the sample size is inadequate or because there isn't, in fact, a suggestion of difference. So, what exactly did you say in the protocol was the intention in the per-protocol?

DR. SABOLINSKI: First, I just would like to make a comment regarding the criteria used. They were a combination of inclusion, exclusion,

and also important factors on case report forms. I would like to turn this over to Dr. Lloyd Fisher who could probably provide a more complete answer to your statistical question.

DR. FLEMING: Lloyd, I really need just two answers. I just need to know exactly what the protocol said per-protocol would be and what did the results show.

DR. FISHER: There is a need to describe to you what went on, and I think you will like it because it is a great teaching example of how not to proceed and I also think the clinicians may like my answer because it shows what happens when you ignore clinical realities.

First, as you can see on this slide, there were a number of people who turned out to be eliminated--I think it was 10 percent or so--because the study stopped early. I mean, it had nothing to do with the protocol; it had to do with the DSMB decision. Then, there were another 10 or 11 percent where they were supposed to have an LV echo within 6 months. It turned out a lot of

people were stable, had had it a little bit earlier and they waived them to go out to 9 months. Then, there were 21 percent in the BiDil group who discontinued their study drug due to adverse events. So, all of these things add up.

But the most important thing, to my mind and I felt good because I brought it up, is that when they looked at the discontinuations--the bottom sentence on this slide, out of 86 deaths there was only 1 included in the per-protocol group. Let me state that again. It seems really astounding. Out of 86, they had 1 left--

DR. FLEMING: Sure. Because patients that died didn't live to a point where they could get the exam required for the per-protocol.

DR. FISHER: No, no, because if you died before the exam you were in there. Of 301 first hospitalizations for heart failure only 39 remained. So, basically, the real data--forgive me, quality of life people, but to me the most important data were eliminated from this analysis. I don't have the figures with me but quality of

life, by the way, was in the correct direction although with very little power.

Heart disease is a continual progressive disease. Heart disease patients get worse, and I know this from being on a lot of DSMBs and how many IIIs went to IVs, and so on and so forth. Well, what do you do if you are treating a heart failure patient, the symptoms increase, the patients are unhappy, and you are supposed to somehow adjust things and they are on blinded medication?

I will give you a little evidence why this might be true but we don't have quite the right data to totally prove it. My theory is you say, well, I don't want to mess around with a lot of different drugs and dose ranging when I don't even know what the patient is getting. So, you take them off the study medication and then you adjust it. Of the people who were permanently discontinued and had a subsequent event, 30 percent of those subsequent events were the day before or the day of the event--in really close proximity to the events. If you went out to I think it was 30

days--I don't have the figures on this slide but I think that accounted actually for about another 6 percent of the people.

My conjecture is, and this is just a conjecture because they didn't ask for reasons for permanent drug discontinuation--my conjecture is that a lot of the rest of the people that had the drug permanently discontinued but, believe it or not, the cardiologists really were able to treat these people and bring them back to an acceptable stable level so you didn't have an event. They weren't hospitalized for it; this was done on an outpatient basis.

DR. FLEMING: This is genuinely interesting to me, what you are saying, because I think you are beautifully illustrating why such analyses like this are highly flawed or have high risk of selection bias. Just in the interest of time, while I would like to hear a lot more about it, this issue, while important, isn't worth a lot of time on the committee.

So, I just wanted to drill down on the

issue--I raised it because of the principle you talked about, Steve. I just want to understand what the results were relative to what was specified in the protocol. So, I just need two simple answers here, exactly what did the protocol say the role of the per-protocol analysis would be? I will even step away from the second answer which is what did the results actually show?

DR. FISHER: It was supposed to be a sensitivity analysis.

DR. FLEMING: So, it was in the protocol purely as a sensitivity analysis?

DR. FISHER: Correct.

DR. FLEMING: All right. I think we can go on.

DR. FISHER: I was going to start saying that in my mind this is a red herring--

DR. FLEMING: Yes.

DR. FISHER: --and it only came up because of the first FDA question. In my opinion, it would have been nice, if there had been enough discussion, to just throw it out.

DR. NISSEN: The only thing I care about is the intent-to-treat analysis, as I think the rest of the committee does. But, you know, we always learn from analyzing studies like this and we try to apply that learning in future trials. I think with your probing here we did learn some lessons. I certainly did and I think we maybe can move on from there.

DR. FLEMING: Or we might say we just had reinforcement of--

DR. NISSEN: What we already knew.

DR. FLEMING: --understood lessons, yes.

DR. NISSEN: I was being polite.

DR. FLEMING: One more question, if I could.

DR. NISSEN: Please, go ahead.

DR. FLEMING: Again coming back to lessons learned in the past, and this relates to missing data, could you show us the slide on the quality of life changes--and I know this is in the briefing document--by outcome and missingness? I want to get at the impact of missingness here.

DR. YANCY: If we can pull up our quality of life slide, it is 0369-12 months?

DR. FLEMING: While you are getting it, the committee can look on page 32 of the briefing document in the medical review. In there, the results are presented for change from baseline and quality of life at 6 months by level, 2, 1, 0, -1, -2, and it appears you did follow your algorithm of assigning worse score to missings, which I would be cautious about. It is conservative on a patient level; it is not conservative on a treatment effect level unless all the missings are in the treatment arm. At least, it does appear that there are more missings in the treatment than control so if there is a bias here, it looks like it is not biased for an exaggerated treatment effect. But when I add up the numbers I don't get the total number of people in the trial. So, it is unclear to me. There appears to be additional missing data. When I add up the numbers on page 32 I get 472 people on BiDil and 500 on placebo, rather than 518 and 532.

DR. YANCY: There are several explanations

and a number of us will comment. If you look at the aggregate patient population, and this is in the FDA review, there is, in fact, an imputed worst case scenario or score for the missing data. If you use the last observation carried forward, which is demonstrated on the slide that was previously up, there are 10 patients in the study in whom we had no quality of life assessment. So, if you look at the column to the furthest right, that is 1040 patients, the last observation carried forward analysis is there and you see that it is statistically significant.

At each of the 3-month increments where we are looking at paired data where data are available at baseline and at 6 months, then we have an additional quality of life assessment which is in the direction for each assessment and misses statistical significance only at 12 months, with a trend at 15 months. I will let Dr. Sabolinski make further comments.

DR. SABOLINSKI: If I may have slide EF-21? What the slide I am bringing up will show

is the percentage of patients that have complete quality of life assessment at each time point. It is the sister slide to what Dr. Yancy just had up.

Dr. Fleming, the table that you were referring to was the categorical scoring system of -2 to +2 where missing values were imputed. All patients should have been included in that tabulation, and if you count them up you will see that they were. In fact, there were 81 patients with missing quality of life. Either they didn't have a baseline, or they had a quality of life without a baseline, or a quality of life that was not within the 6-month period of time.

DR. NISSEN: Tom, I would guess here that this is an artifact related to the early termination because the study terminated before patients reached the 6-month quality of life time and, therefore, they would have no assessment. They are not truly missing. Don't you think, Bob, that is what is going on here?

DR. TEMPLE: I am not sure I am following all of this, but one of the things that happened is

that the imputation rule, if you didn't have a final value, was to give you a worst case analysis. So, the fact that a lot of people didn't finish the study as planned gave a lot of people a worst case analysis which sort of equalized everything. So, the analysis that did that, we thought--we were talking about this a couple of days ago--was probably flawed and we have actually asked for further analyses month by month.

DR. NISSEN: I am going to guess that what is going on here is that this is an artifact of the early termination, don't you think, Tom?

DR. FLEMING: It may be. I think it is clear to me what is happening here now that you showed that slide. You talked about an LOCF method. So, were you using LOCF here or were you actually imputing worst case if you didn't assess? On page 32 it looks like if it is missing you are imputing -2.

DR. SABOLINSKI: There are two different analyses for two different purposes. Again, for the primary efficacy endpoint, looking at the

quality of life as a component of the composite, we used a categorical scoring system where it was prespecified in the protocol that a -2 would be imputed for missing data. That calculation is provided both by the sponsor and in FDA's briefing book. What that was intended to do was to show the relative weight of quality of life in the overall composite score. What was prespecified in the protocol was a secondary analysis looking at quality of life overall in the study at the time points that were shown in Dr. Yancy's slide. Those data were done by paired quality of life scores. Those patients, at each time point that had a baseline and a quality of life score at each prespecified time point, were included.

We also prespecified in the statistical plan that they would be compared by a 2-sample t-test. We used the Minnesota Living with Heart Failure scores in this case which is a continuous scoring system, and we took the difference between the time point and baseline and did the 2-sample t-test.

DR. FLEMING: So, in your analysis of the quality of life change at 6 months, I believe you reported the p value--well, you tell me, what is your reported p value for that component alone in the analysis of quality of life?

DR. YANCY: Can we bring that slide up?

DR. SABOLINSKI: It is 0.011.

DR. FLEMING: And in that analysis the specific way you handled missing observations was LOCF or imputing worst value?

DR. SABOLINSKI: Neither. It was done by paired tests where you had baseline and you had a score at the 6-month time point, as defined in the protocol.

DR. FLEMING: If that score wasn't available?

DR. SABOLINSKI: Then the patient was not included in that scoring system.

DR. FLEMING: I see, so you deleted that patient.

DR. SABOLINSKI: That is correct. And I did show the slide, I think it is EF-21, which

shows the percentage of patients at each time point that had paired scores.

DR. FLEMING: Can we presume that the number that were deleted then corresponds to this number, this 81 total? Is that right, 46 and 35 or 81?

DR. SABOLINSKI: At the 6-month time point.

DR. SACKNER-BERNSTEIN: Could I just ask for a clarification about that?

DR. NISSEN: Please.

DR. SACKNER-BERNSTEIN: So, are you saying that 46 and 35 patients--that is all that didn't make it to the 6-month time point? Because that doesn't make sense. That is inconsistent with the rest of the data.

DR. SABOLINSKI: There are some patients that did die and those patients certainly were not included and did not have the score imputed.

DR. SACKNER-BERNSTEIN: Looking at these documents and also the design paper that was published, what I don't see any evidence for is the

strategy that I think you have used based on the fact that the study was stopped early. It says all the way throughout that the quality of life components will be based on 6-month outcome and that anybody who doesn't have data would get the worst rank.

DR. SABOLINSKI: As a secondary endpoint. Minnesota Living with Heart Failure scores were used and the difference was taken from the point in time measured and compared to baseline, and between-group differences were compared by the 2-sample t-test.

DR. SACKNER-BERNSTEIN: Okay. But for the primary analysis--

DR. SABOLINSKI: For the primary analysis the component was used using the categorical -2 to +2 scoring system.

DR. SACKNER-BERNSTEIN: Now, for people in the trial as far as the primary outcome is concerned who did not get to 6 months, my impression is that what you did was you used the 3-month quality of life outcome; you used that as

the last observation carried forward and put that into the primary endpoint analysis. Whereas, if that is correct, if that is what you did, I don't find documentation that took into account that that is what you would do if the study ended early.

DR. SABOLINSKI: What we did in the statistical plan, once the study was stopped early and prior to unblinding, we stated that we would use the last score after baseline and that was carried forward to the 6-month period of time.

DR. SACKNER-BERNSTEIN: What was the result of the study without that change so that anybody who didn't have a 6-month quality of life outcome gets assigned a -2? What would the score look like then?

DR. SABOLINSKI: We haven't done that analysis.

DR. SACKNER-BERNSTEIN: Well, that was the prespecified analysis--

DR. SABOLINSKI: Actually, the prespecified analysis in the final statistical plan was to use the post baseline score and carry that

forward.

DR. SACKNER-BERNSTEIN: I don't know whether this is a concern or not but it sounds like you created a statistical plan alteration after you knew that there was some outcome in the trial--

DR. FISHER: May I make a quick comment?

DR. NISSEN: Please.

DR. FISHER: In order to get different types of data, two parts which were really time-to-event data, survival and first hospitalization, and then you have this quality of life questionnaire score--in order to combine those in some way things were categorized. Statisticians, at least non-epidemiologist statisticians, do not like to throw away information so you will notice that death for the composite is just yes/no, it occurred or not. Yes, we have presented survival curves with a log-rank which was the preplanned analysis for the survival data because you don't throw away all that information.

Similarly, there was categorization of the

quality of life questionnaire which throws away a lot of data. It is a continuous scale, as you heard, from 0-105 and it was grouped into those 5 intervals in order to get this composite. But for a statistician, it sort of offends your soul to throw away information that is relevant to the decision you are making. So, before unblinding the plan for the analysis of those data was the 2-sample t-test, not to work with the categorization that had been used for the composite score.

Similarly, for first hospitalization it was given a score as a component of the composite but the analysis of those data, when analyzed alone where you have the time-to event-data, is the log-rank statistic, time to event.

DR. SACKNER-BERNSTEIN: I am not sure I would react any differently if I saw the data because I think that doing the analysis based on the way it was initially prespecified, as opposed to the finally prespecified version, would necessarily change how I would interpret things

because, of course, the study has a bias if you are going to automatically assign anybody who doesn't get a 6-month assessment, because of the early termination, the worst rank. But it is important to have the transparency, as Steve was talking about before. If it even went just kind of in the same direction but was nowhere near a statistical p value, that would be very reassuring to me as opposed to the change to a last observation carried forward even though it was prespecified in the final analysis plan.

DR. ARCHAMBAULT: It was always prespecified in the protocol--always prespecified in the protocol that if the 6 months quality of life were not available the 3-month quality of life would be used, in other words, an LOCF. That was prespecified in the protocol.

DR. SACKNER-BERNSTEIN: But it is not in the documents.

DR. ARCHAMBAULT: The statistical analysis plan, which was finalized prior to the study even ending, said that we would use LOCF for any

subsequent to baseline quality of life. Can we pull that slide up?

DR. NISSEN: I want to make sure that we understand it. I am getting confused here and I am not a statistician, I am just a poor cardiologist. But what I think we are trying to get at is after you knew that the data safety and monitoring board had stopped the trial were the statistical methods then modified and there was a new statistical plan? If so, how were they modified? What I think Jonathan, Tom and others are getting uncomfortable with is that once you know the outcome of the trial that has been stopped, if you then change the methods you are using for analysis, that has some implications. So, what I want to know is did you amend the SAP after you knew that the trial had been stopped by the data safety and monitoring board?

DR. ARCHAMBAULT: No. The answer is no.

DR. NISSEN: Good.

DR. ARCHAMBAULT: May I see that slide? I would like to point something out on that slide,

EF-22? Thank you. The top portion shows the second bar, the 6-month bar and Dr. Yancy's slide that showed each of the 3-month quality of life. The top one is just paired data so we have 369 patients and 371 patients during the change from baseline in the actual Minnesota quality of life score. So, there was a mean 7-point reduction in the BiDil group and a mean 3-point reduction in the placebo group.

The bottom half of the slide, which I am sorry is kind of hard to see from here, shows the 6-month assessment using LOCF but again using the entire Minnesota Living with Heart Failure questionnaire and LOCF was implemented this way: If a patient did not have a baseline, and there were 10 patients who did not have baselines, they were not included because there was nothing to carry forward and how could you do a change if you didn't have a baseline?

Perhaps we could have gone back and done next observation carried back, or something like that, but there were 10 patients without baseline.

So, you see that the sample size is 512 and 528 rather than the 518 and 532. There were 6 patients on BiDil and 4 patients in the placebo group who did not have a baseline quality of life performed. All of the remaining patients are included in there.

So, what did we do for patients who through the 6-month time point had a baseline but no subsequent value? In doing the last observation carried forward, we assumed that the score was the same so the change in score for each of those patients would be a score of zero. There were a total of 81 imputed scores for the component in the composite score. So, that left 71 scores of zero that were in there among the patients in the BiDil and placebo group, and with LOCF we do, in fact, have still nearly an 8-point reduction and a 3.4 reduction, and the p value is more significant than just at this particular time point.

So, by including more patients using LOCF, which we don't think biases things here because there were more zeroes that were stuck in on the

BiDil side, in fact we still have statistically significant results. We are hoping that this addresses the issue as to whether or not quality of life is reasonably well supported at 6 months either with only the patients that have paired data at that point or with an appropriate technique for carrying forward observations.

DR. FLEMING: Just a very quick comment on that, it obviously is an intrinsically complicated situation when you are trying to assess effects on quality of life, and it is admirable that you are, and what you have is the ability to assess about 73 percent of the people--that is what this figure says, 73 percent or so of the total sample size is actually assessed at 6 months. So, you are having to do something about that 27 percent, and there are some traditional approaches, each of which or all of which are very dissatisfying: Wworst case, very dissatisfying because it is clearly biased. It may not be biased in favor or against treatment. It all depends on the nature of who it is that is missing. Another is LOCF. Is that rational? Most

of the time LOCF is not rational.

So, my point here is the placebo arm has a decline over time so natural history is a decline. The best way to make a treatment look good, therefore, is to stop following people on the treatment arm soon and impute LOCF. That means you are imputing no change in a scenario where people are declining. So, LOCF, to my way of thinking, is never a good thing but it is a terrible thing in a setting where you are trying to get stabilization where there is decline.

Now, I am not saying that you are getting tremendous bias here from that because there is LOCF-ing in both arms, but there is a little more LOCF-ing in the BiDil arm. The bottom line is a lot of this missingness is due to deaths as well, and we shouldn't be trying to factor that out. Death should be included, given some type of worst score.

Mr. Chairman, I am guessing that we probably don't need to discuss at a lot greater length, except to say I think this missingness is

something that is worth understanding and realizing that it provides some complexity in being able to interpret the signal.

DR. SABOLINSKI: If I may just come to a point and clarify, as we have all seen, the primary composite score was made up of death, quality of life and first hospitalization due to heart failure. Each had a categorical scoring system. That was used to determine the primary composite score. For secondary endpoints, statistical methods were used and prespecified so the Kaplan-Meier curve with the log-rank test was prespecified because it used all the data. For first hospitalization due to heart failure, Kaplan-Meier with the log-rank test used all the data. For quality of life, we said that it would be Minnesota Living with Heart Failure questionnaire score change from baseline and compared between groups overall in the study.

So, I think it is important to distinguish what was a contribution to the primary efficacy endpoint and how that was measured, and how quality

of life was intended to be measured throughout the trial.

DR. FLEMING: We understand that, and I think Lloyd Fisher clearly made those points earlier. The essence, the bottom line here is for any measure you want to come very close to achieving complete follow-up. If I think across disease areas, cardiologists do extremely well in being able to achieve that goal overall. Some measures are harder to follow uniformly than others. Fortunately, for your heart failure hospitalization you only have 24 people missing--not perfect but a relatively small number. Here we have 81 and I am not even sure if it is not more than that in terms of the way it looks here. It looks like about a third or 27 percent that weren't fully assessed at 6 months.

So, the bottom line here is any time you are not making a complete assessment of your endpoint there are significant risks of bias and, hence, complications in interpreting the results, especially when you get up to 27 percent that

aren't assessed at 6 months.

DR. SABOLINSKI: Well, actually, many patients didn't make it to 6 months in order to be assessed, and that was due to the early stopping of the trial. But I would like to make two points. First, the placebo group does not deteriorate over time. Second, with the increased death rate in the placebo group, what you are left with is basically sicker patients or potentially sicker patients in the BiDil group and that is a bias against the BiDil group for assessing quality of life.

DR. NISSEN: We understand that. We understand it very well. I am almost ready to quit on this one but I have one more thing I just have to point out. Put up CE-15 again because I do think that, again, we are trying to learn what we can and I think there are some things to learn.

Now, earlier on I asked the question why did you pick the 6-month time to assess quality of life. I have to point out to you what would have happened if you had arbitrarily picked 12 months to assess quality of life. You see that the p value

at 12 months is 0.13. Now, trends are all in the same direction and that doesn't necessarily mean that there is not an effect, but it is arbitrary and when you make an arbitrary selection like that, you know, one could argue that you got lucky here that you picked the right time frame because there is some variance here from time to time.

So, a learning here might be--and I don't know whether there are good statistical measures for doing this, but I am kind of interested in understanding what the quality of life is over the course of time the patient is on treatment. You see that here. If you look at it, there is a consistent effect but there is some variance at single time points. I don't know whether the agency comes up against this in other trials, but it seems to me that a more useful clinical measure, if there are good statistical ways to report it, is almost as an area under the curve kind of measurement where you are looking at the total quality of life over the course of the trial. Bob, did you want to comment?

DR. TEMPLE: Well, only that I am sure that 6 months represented some attempt to be persuasive that it was long enough to matter and not so long that you have a lot of dropouts. So, by the time you get to 12 months there are fewer patients; the point estimates don't look quite as good but later on they do look good. Who know what those things mean? And, we tend to be inclined to believe the one that, for better or worse, they picked.

But the whole question of how to do quality of life--I mean, millions of people do quality of life assessments and hardly any of them are ever persuasive enough to get into the label. This is a very useful discussion. It is not at all clear how to do it.

Even when Tom says he doesn't like LOCF, it is very important for us to know what he does like--I assume some modeling approach that takes into account that people weren't there and tracks the fact that they are declining. We worry about that in many other areas. In all of the

psychotropic drugs LOCF is standard operating procedure for depression trials and everybody hates it. We don't quite know what another way of doing it would give you and whether it would be different but we are actively exploring those things. We have whole conferences on how bad LOCF is.

DR. NISSEN: I am just commenting on some discomfort in picking an arbitrary point in time and saying that we are going to look to how you feel at exactly 6 months into treatment and that is quality of life. If you think about it from the point of view of a patient, which is obviously what we have to do, what they like to know is that every day, every month, every week, every year that they feel better is contributing to their quality of life. As a clinician, I would like to understand for any drug what the effect is on the overall quality of life during the course of therapy. If there are ways to do that, I think that is very important.

DR. TEMPLE: It is all true. Quality of life is always, let's say, muted if there is a

survival benefit because only the better people manage to stay in and give you a score. So, you look at all that and you say that is despite the fact that the dead people aren't contributing. It is also tempting to give people who die some score that you carry further but then, you know, you argue about how big the score should be. I don't know a simple way. Tell me if this is crazy, but what is at least slightly reassuring is that all the bars sort of look similar over time, even though that is a different population every time and, of course, people who have dropped because they were hospitalized repeatedly are gone. People have dropped out because they are dead or gone. And it still seems to be there. I guess I would say that is reassuring but that is not a very quantitative statement.

DR. NISSEN: Actually, it is very reassuring and that is one of the first things I noted. But the day may come when we sit around this table, or our successors do, and we see a situation where the quality of life is better for

the drug at 6 months and worse at 12 months but the prespecified was 6 months. So you get all the points for prespecifying it but then you ask, as a clinician, what did we do of the patient? We made them better for 6 months and we made them worse for 6 months and that is not meaningful. So, you have to put on the hat--you know, it is not a statistical question; it is a clinical question. So, what reassures me here is when I see consistent differences, regardless of the p values. For each assessment the bar for BiDil is better than the bar for placebo. Now, it is not a statistical argument, it is a clinical one, but it does help me.

DR. TEMPLE: We would think of those as supplementing the fact that they got lucky and succeeded on the thing they prespecified, which is sort of critical. If that had not been prespecified and you just looked at bars we would be more nervous.

DR. NISSEN: Well, you are going to have to write the label here and one of the questions,

obviously, for all of us to think about is, is improvement in quality of life--did they achieve significance for labeling purposes? I think we would like to discuss that.

DR. TEERLINK: This is just a point of clarification on how the primary endpoint was calculated. At 3 months you had an increase in your score by 10 or an improvement in your Living with Heat Failure score by 10, and then died at 4 months. With the technique that was employed for the primary endpoint analysis did you get an improved score? What was your score for your quality of life if you died at 4 months but had an improved quality of life for 3 months for your 6-month endpoint?

DR. SABOLINSKI: For the primary efficacy endpoint, since it was specified that last observation carried forward would be implemented, the improved score--

DR. TEERLINK: So, if you died at 6 months--

DR. SABOLINSKI: Negative 3 for death--

DR. TEERLINK: Negative 3 for death and then--

DR. SABOLINSKI: Then whatever the categorized version. If you said it was 10 points then, you know, I am not sure whether that was the +2 break point.

DR. TEERLINK: The +2 break point.

DR. SABOLINSKI: So, it would be +2. So, it would be -1 and then one would have to know whether or not one was hospitalized or not.

DR. SACKNER-BERNSTEIN: I need to just follow-up on the primary endpoint analysis because once again I am confused by what I am hearing and what I am reading. In the NitroMed briefing document it points out that if a post baseline quality of life assessment was not available the patient was assigned a worst score of -2. Yet, I think from the last question I asked there were some 81 patients, or something along those lines, who were assigned a rank of zero if they had a baseline but they had nothing else after that. That is what I thought was said.

DR. SABOLINSKI: No, for the component those 81 patients who had missing data were assigned a negative 2.

DR. SACKNER-BERNSTEIN: Good. The second part of that is that the FDA document, page 26, 5.1.10 says post hoc changes, after the termination of the study the sponsor requested the addition of analyses, termed sensitivity analyses, in which missing data were to be handled differently than originally planned. This is what I was getting at before when I was trying to understand whether there was a change in the statistical plan for the primary endpoint. My impression from these documents is that anyone who didn't make it to the 6-month visit to have a quality of life assessment performed should have had a -2 included in their score in the primary analysis in the intent-to-treat population. I need to be reassured that that is a mis-impression. With it is or isn't, I still would like to see the results with those patients who didn't have a 6-month score given a -2.

DR. SABOLINSKI: First, it is a mis-impression that the statistical plan and the prespecified primary efficacy endpoint did not change. What was done post hoc was the definition of sensitivity analyses, and these analyses imputed scored in various ways. The three analyses were shown. Each of them is statistically significant. The first is with no heart failure hospitalization imputation. The second is with last known quality of life carried forward. And, the last is the analysis done on only those patients that would have had the opportunity to have completed 3 months, that is, when the study was stopped. This was looked at as being perhaps the purest population and the one that would have had all the data that was intended when the protocol was written. But these are sensitivity analyses only and were intended to show the consistency of results when compared to the prespecified primary endpoint in the study. The primary endpoint in the study did not change; was prespecified and was the basis for the analysis that Dr. Yancy showed as his

first data slide.

DR. NISSEN: All right, are there any more committee questions? Jonathan?

DR. SACKNER-BERNSTEIN: Do we want to talk about some of the safety data now or do you want to get through the particulars and discuss that later?

DR. NISSEN: Maybe we can do that later. Let's see if we can't get through the final presentation. We are obviously running a little slower than had been originally anticipated but that never surprises me. So, let's hear the conclusions. I think we are going to have Dr. Packer, who has spent a lot of time with this committee over the years. Milton, tell us what you think.

Conclusions: From V-HeFT I to A-HeFT

DR. PACKER: Thank you very much, Steve. I would like to briefly highlight the key lines of evidence supporting the approval of BiDil for the treatment of heart failure and provide an opportunity to answer any questions about the trials and the specific drug. I want to

specifically focus on the issue of strength of evidence that Dr. Hiatt and Dr. Fleming addressed.

The evidence supporting the approval of BiDil is derived, as you have heard, from three multicenter, controlled clinical trials that have evaluated the effects of isosorbide dinitrate and hydralazine in chronic heart failure. The two V-HeFT trials enrolled patients who had primarily mild to moderate heart failure, who were treated only with digitalis and diuretics. In contrast, the A-HeFT trial enrolled patients who had primarily moderate to severe heart failure, who were treated with ACE inhibitors, beta-blockers, aldosterone antagonists in addition to digitalis and diuretics. Women were only enrolled in the A-HeFT trial and white patients were only enrolled in the V-HeFT trials.

I want to note that these trials do not contribute equally to the assessment of the efficacy and safety of BiDil for the treatment of heart failure. As emphasized in the briefing document and throughout today's presentation, the

key trial here is the A-HeFT study which was the only one in the three trials carried out according to currently accepted standards for protocol development, hypothesis testing and data collection. The V-HeFT trials, however, played a key role in generating hypotheses that were tested in the A-HeFT trial, but they also provide an opportunity to confirm the findings in the A-HeFT trial.

So, let's remind ourselves what the prespecified primary endpoint in A-HeFT was, as was described in the original protocol, agreed upon with the FDA before the start of study, was a clinical composite score with three components: death due to any cause, hospitalization for heart failure and quality of life.

Now, you have already seen that BiDil was superior to placebo on this prespecified primary endpoint according to the prespecified primary analysis, and this superiority is a key element in supporting the approval of BiDil for the treatment of heart failure in black patients.

Now, as the FDA and as this committee has emphasized on numerous occasions, whenever you have a composite endpoint it is important to make sure that each component contributes importantly to the success of the endpoint. You have already seen from Dr. Yancy's presentation the effect of BiDil on death, on the left; on hospitalization for heart failure, in the middle; and quality of life of life. And each of these contributed importantly and separately to the primary endpoint's success. It is important to notice that the original protocol also specified that the effect of BiDil on each of these three components was to be analyzed individually, and the protocol gave weight to these individual analyses by designating them as the leading secondary endpoints in the trial. The individual analyses of each component not only allow us to confirm the independent contribution of each component to the success of the A-HeFT trial, but they also allow us to look for confirmation of similar benefits in the black patients enrolled in the V-HeFT trials.

Let's first look at mortality. In the black patients enrolled in A-HeFT BiDil reduced the risk of death by 43 percent. You can see the p value of 0.012. You will see, just out of interest, that the curves began to diverge perhaps at around 6 months in this study.

This finding is strikingly similar to the effects of isosorbide dinitrate and hydralazine in the V-HeFT trial which reduced the risk of death in black patients by 47 percent. The p is 0.04. Again, the curves began to diverge at about 6 months. Obviously, in this study this was the hypothesis generating observation confirmed in A-HeFT but, in fact, when we take A-HeFT and we look for confirmation the subgroup analysis in V-HeFT does provide for consistency across the trials.

The same principle applies to hospitalization. In the black patients enrolled in the A-HeFT trial BiDil reduced the risk of a heart failure hospitalization by 39 percent--very small p value. This is in spite of the problem and issue

of competing risk with mortality. You will see that the curves separate early and maintain their separation. Remember, the follow-up period here is 18 months.

This finding is concordant with the findings on hospitalization in the V-HeFT trials. Remember, the issue of hospitalization in V-HeFT--non-adjudicated, nor recorded at the time of event. But despite these important limitations, you still can construct time-to-event analyses. They are far more imprecise for V-HeFT than they are for A-HeFT. If you look specifically at the first 18 months, you can see that treatment in these two trials with isosorbide dinitrate and hydralazine in black patients was associated with a lower risk of heart failure hospitalizations when compared with placebo in V-HeFT I, and when compared with enalapril in V-HeFT II. Remember that enalapril has been shown to reduce the risk of a heart failure hospitalization when compared with placebo.

Finally, let's look at quality of life.

The committee has already gone through this to a substantial degree. The prespecified time point for analysis here was 6 months but all other time points were prespecified as secondary analyses in the trial. There is consistency across the effect and that is true in the first 3-6 months, which is the conventional time period for looking at quality of life in heart failure trials. But it is also true later in the trial when the issue of competing risk of mortality becomes an issue, and because there are higher risk patients in the BiDil group because of their lower risk of death, the analyses towards the end of the trial are actually biased against BiDil. In spite of that, the separation between BiDil and placebo is maintained.

It is interesting--Jay would be able to comment on this--there aren't too many trials in the history of heart failure where we have seen this kind of consistency in terms of benefit of quality of life across all time points across the prespecified duration of the study.

These findings on quality of life in

A-HeFT are concordant with those in V-HeFT II. Let me emphasize that the reason I picked V-HeFT II is that it is the only one of the V-HeFT trials that actually assessed quality of life. Quality of life wasn't assessed in V-HeFT I. In the black patients enrolled in V-HeFT II the combination of isosorbide dinitrate and hydralazine had effects on quality of life that were at least as favorable as those produced by enalapril. Remember, lower scores are better quality of life. You can see that isosorbide dinitrate/ hydralazine group in yellow and the ACE inhibitor enalapril group in orange. Remember, this is a comparison versus an active control which has favorable effects on quality of life.

Thus, if one looks at the totality of available data, there is really very good concordance within and across the controlled clinical trials with BiDil. These trials enrolled black patients with mild to moderate heart failure, treated with digitalis and diuretics, as well as black patients with moderate to severe heart

failure, who were also treated with ACE inhibitors, beta-blockers and aldosterone antagonists. The concordance of the benefits across three clinically relevant endpoints, all prespecified as the components of the primary endpoint in A-HeFT, would support the proposed indication which states: BiDil is indicated for the treatment of heart failure as an adjunct to standard therapy in black patients to improve survival, prolong time to hospitalization for heart failure, and improve quality of life.

I will be happy to take any questions.

Questions from the Committee

DR. NISSEN: Questions for Dr. Packer? I think we will also try in the time we have before lunch to address any additional questions related to Dr. Yancy's presentation. I think Jonathan, you had some questions about AEs. So, questions for Dr. Packer?

DR. HIATT: Thanks. I really appreciate the overview. Two questions, one is on the strength of evidence. We are left with a p value around 0.1, not 0.00125. One wonders, had the DSMB

not stopped the study, whether group separation might have continued to broaden. Certainly, we respect that decision and understand the clinically motivated reasons to do that. So, you might just comment on the strength of that p value.

My second question, which we haven't really talked about and I would like your impression about this, is when you look at combination products usually you have an A and a B and each has to beat placebo, and then A plus B has to be better than A or B. We don't have that here. What are your thoughts about that? It is a fixed dose combination. We are kind of stuck with not a lot of dose-ranging information here, another sort of missing piece that I look for in drug approval. So, those are my two big questions.

DR. PACKER: Let me take the first question. If you look only at the primary analysis of the primary prespecified endpoint, I agree with you. We are looking at the strength of evidence around 0.01. You have to feel better about it than just 0.01 because you have now all three components

contributing. Each of those components actually reaches statistical significance in its own right in spite of the fact that two of the components are biased because of the competing risk of mortality. And, you look then at V-HeFT. Remember, V-HeFT I and V-HeFT II were hypothesis generating for mortality but you have not only mortality data in that study which is meaningful. But you have hospitalization data; it has some issues. You have quality of life data.

So, you can not only look at the primary component, 0.01, you can look at the consistency of its components. You can look at the consistency of those components within A-HeFT over time. It is durable. Then you look at the consistency of each of those components against V-HeFT and you have consistency. There you get up to a strength of evidence which is the conventional standard of two trials. It is not a mathematical way of getting there; it is an intuitive way of getting there.

DR. HIATT: The second question was the different components. We are stuck with fixed

doses. We don't know really what drove these results. Your impression about the weakness inherent.

DR. PACKER: The real problem is that we just don't have a whole lot of evidence saying that isosorbide dinitrate alone or hydralazine alone is effective in the treatment of heart failure. The trials summarized in the briefing document are not awful trials but they never have shown the between-group difference in favor of either drug used as monotherapy. So, there is no way that anybody can say that either of these drugs works as monotherapy.

Now, the trial here used a fixed dose combination. The only thing that can be said is what you see is the result of use of the fixed dose combination. It is impossible to tease out what the contribution might be of each component.

DR. HIATT: Bob, other drug development programs require this. Why is that not an issue here?

DR. TEMPLE: Good question. There is

actually a regulation called 300.50 that says that two drugs may be combined in a fixed combination when each makes a contribution to the claimed effect. It doesn't say how you have to demonstrate that but, as you know from your question, the usual way you do that is to compare A/B with A and B and be better than either.

We have grappled with this over the years and considered the kind of data that might go into a conclusion that the combination is better than the components. You have heard from Milton and from the presentation that there are some arguments for explaining why neither one alone would do the job but that is not like having a trial showing it.

We have been confronted with the potential for this in the past and, in the form of an old memo from 20 years ago that Marian Finkel wrote, and based on some thinking now, we also worried about the situation where you have data that would make it uncomfortable, to say the least, to explore which of the two components makes this contribution. In other words, if you wanted to

conclude now that they have made a mistake by doing the combination and you wanted to find out which one contributes, you would have to do a trial to show that people on one of those drugs die. There might be circumstances in which you did that if one of the drugs was very toxic, or something like that, but one of the things that you all have to help us grapple with is whether you really want to tell people that they have to do a study in which you will discover which of the two components saves your life by showing that people who don't get that component die more frequently or do very badly more frequently.

But we have said that some judgment applies to this in the past. We are actually reworking our combination policy rule, and we still think that you have to be reasonable on some of these things if you have an important endpoint. This wouldn't apply to minor symptomatic benefit, but if it is a major endpoint you have to ask whether you can still do the study in question.

DR. NISSEN: I wanted to ask you, you

know, we talked about the sort of evidence question and the p value, and maybe Tom is going to comment on this as well, but there is obviously an adjustment for the interim analyses. You know, in terms of how that is done in this trial, how was that done? I need to understand that a little bit better. Can anybody explain that to me?

DR. D'AGOSTINO: Ralph D'Agostino, Wake Forest University. I performed the interim analysis that provided the sample size calculation. Are you asking a question about that or the ultimate, final p values?

DR. NISSEN: Yes, the adjustment for the interim looks.

DR. D'AGOSTINO: Tad can probably speak better to that.

DR. ARCHAMBAULT: We adopted an O'Brien-Fleming boundary with two interim looks and a final. So, the total number of looks was scheduled to be three. They were scheduled to be at 25 percent, 50 percent of the original patient accrual which would be 300 patients at six months,

and then a final look. With the sample size re-estimation that was done at the second one--well, first of all, the original p value for the third look was to be 0.48 using the O'Brien-Fleming boundary. So, the overall probability or type-one error would be bounded above by 0.05. So, the last look was to be done at 0.48.

By doing the sample size re-estimation we could keep the same nominal p value but there was an adjustment to be made to the statistic according to the Cui, Hung and Wang method, and that is what we did.

DR. NISSEN: So, the adjusted p value is not 0.01.

DR. HIATT: No, it is not. In fact, if you look on page 32 of our document, there are three p values. There is unadjusted, 0.011; sponsor's adjusted, 0.016; FDA's adjusted, 0.021.

DR. NISSEN: Yes. I kind of knew that but I wanted to make sure we kind of got that out on the table, that the strength of evidence is not

0.01; it is really 0.02. Maybe that is not important but we do these things for a reason and I wanted to make sure we all understood that. So, the final adjusted p value is 0.021.

DR. ARCHAMBAULT: Well, there is some open discussion relative to that. We feel that 0.016 that we provided is appropriate. Dr. Hung feels that 0.021 is appropriate. And a third method of calculating the statistic, which includes all of the interim analysis to patient data in the first portion of the statistic, is the one that provides the 0.011.

DR. NISSEN: Tom Fleming, what is the right method for doing this?

DR. FLEMING: Well, there is a lot more to say. We will discuss it when we get to the issues later on. But I don't accept any of those three.

[Laughter]

DR. NISSEN: Having worked with you for five years, I kind of knew that. I want to be enlightened. I mean, we have to do due diligence here so I am looking for help here. Bob?

DR. TEMPLE: I just want to mention one other regulatory issue that has been touched on. As you know, we agreed at a meeting with the company that a study entirely in a black population would be okay. It is worth considering that whole issue and then I will tell you why we might have thought that.

There is tremendous interest in individualization of therapy now and whenever you try to identify the population in which a drug is going to be effective there is always the question of how much information you need about what you might call off population, the group you are not going to study. It is hard to know what is enough. If a drug works less well in that population, providing convincing evidence that it doesn't work requires a massive study. So, how much do you need? I introduce that to say that we have not worked that out.

There are other examples. Lotronex was approved for irritable bowel syndrome in women and there was some evidence that it didn't work very

well in men. But I think if you asked a lot of gastroenterologists, they would say hm, I'm not so sure it doesn't work in men. So, that is an ongoing debate.

Milton didn't emphasize at all the results of the V-HeFT I and II in the non-black population and, you know, that may be a point of massive sensitivity and everybody wants to stay away from it. But one of the features of the V-HeFT I and II is that in both of those studies the white subset of the population, or the self-declared white subset of the population is larger than the black population and it certainly doesn't look like there is much going on. That is what we thought and that is why we agreed that it was reasonable to study primarily the black population. But that is a point that ought to be at least somewhat discussed.

We don't have a firm policy yet, I don't think, on what do you do if someone just sets out and says okay, I'm going to work this up in this population and I don't really care about the other populations because I want to do this. We have

never said that is out of the question but it makes you uncomfortable. Obviously, you want benefits to go to all people. So, it is very clear we expect at least some kind of evidence in the other population and one of the questions here is how persuasive is that, and I don't want to hide it, we obviously thought it was at least convincing enough so that we would accept a trial in the black population. I just want to put that on the table. Milton didn't really address it when he summarized V-HeFT I and II and perhaps it ought to be summarized.

DR. NISSEN: Bob, we have sat around this table and complained many times about getting trials where we don't have enough women and we don't have enough minorities to come to some conclusions, and I think that one of the prices that you have to pay for getting that information is to accept that this kind of exploration is desirable, that having more information about groups that are going to respond, particularly when they are groups that are under-represented in a lot

of clinical trials, is a public policy advantage in having that information.

So, we will discuss this and I will opine about this in the afternoon about whether you give points for the fact that you have done a study in a group that we need more information about; that suffers a lot of burdens from this disease. I will offer my own personal opinion about that, but I think it is relevant, as a society, that we talk about these issues, and probably this panel is a good thing, that we actually think that through and talk out loud about it with a lot of people in attendance. So, I will have more to say about that later.

DR. SACKNER-BERNSTEIN: In the V-HeFT I and V-HeFT II studies were there any Asians or Hispanics enrolled?

DR. PACKER: There were very few I think.

DR. ARCHAMBAULT: Excuse me, we did post hoc sub-looks at it. For race in both of them, both V-HeFT I and V-HeFT II, the classifications were black, white and other. "Other" was not

specified but I think in one of the trials it was 9 patients and in one of the others it was a very small number, perhaps 10-12.

DR. SACKNER-BERNSTEIN: I just wonder where those ethnic groups would have been fit in if they would have participated in the trial. There were very small numbers according to the numbers I have seen and that may be relevant to try to identify how to apply that data as we try to integrate that.

I am also curious a little bit about AEs. Maybe this should go to Clyde. I note that there are some adverse experiences reported--I am talking in general, not necessarily the serious AEs--that are reported across the different trials with hydralazine nitrates at somewhat different frequencies. I look at headache and dizziness and there is not that much of a difference between the side effects in African Americans in the V-HeFT studies versus A-HeFT. For example, headache was in 72 percent in V-HeFT studies; in A-HeFT it was 49.5 percent. Both of those were a little higher

than in the placebo. Dizziness, 67 percent; in A-HeFT it was only 32 percent.

That may go to differences in the populations and the assessment, etc. I think it is interesting that when you have a group of patients in A-HeFT who were treated with other drugs that tend to lower blood pressure, even though the exposure to the study drug is shorter because the follow-up is shorter, dizziness is less frequent. I am finding that somewhat surprising.

Then, what really struck me in a drug that is known to produce a lupus-like syndrome in a small proportion of patients as you look at some of the things that may be reflections of lupus, for example arthralgia--now, I can't say that arthralgia is lupus and I would not want to have anybody infer from this question that I think that way but that is the closest AE that I can find on the list to that syndrome, and in the V-HeFT studies the African Americans on hydralazine and nitrates had arthralgia--65 percent of the patients had arthralgias and in A-HeFT 1.5 percent had

arthralgias.

The reason that stands out in my mind so much is that here we have a drug where the FDA approved product insert for its hypertensive indications suggests pretty strongly that there should be blood tests looking for the serologic evidence of a lupus-like syndrome. There were no such blood draws done in A-HeFT. I think it is very interesting that there is lack of evidence for a concern in this population in this disease state with this drug, largely because it wasn't looked for.

I also would point out that the literature gives me the impression that if a lupus-like syndrome was going to develop with hydralazine it tends to develop after six months; tends to be dose related; and there is at least one report, without terrific data but the best I could find, that women could be up to four times at higher risk for the syndrome.

So, I would like to be reassured that in a study that was stopped early, certainly with great

basis for doing so, it is the right thing for us to do to be satisfied with no information about whether this drug combination that could pose the risk might actually be exposing people to risk.

DR. PACKER: Jonathan, let me take part of that question and then I will hand the rest of it over to Clyde. There are two real big differences between the AE data in A-HeFT and the AE data in V-HeFT. One big difference is the duration of follow-up. The duration of follow-up in V-HeFT is meaningfully longer than in A-HeFT, and that accounts I think almost entirely for the greater frequency of reports in V-HeFT than in A-HeFT. People could report anything for up to five years, whereas in A-HeFT they could report only up to 18 months.

Second, if you remember, in A-HeFT AEs were recorded according to what might be called current policy. They are all spontaneous reports, patients to investigators. In V-HeFT there was a checklist of AEs that the investigator filled out, which makes it sometimes a little bit hard to

compare adverse effects across the two. Of course, there is this big category of AEs in V-HeFT called "other" which we have very little further detail on.

Lupus, however, was a big focus of safety in V-HeFT. They really drilled down on it. Remember, V-HeFT had a greater capacity to look at it because of duration. You could argue they had less capacity because they eliminated women and women are four times more frequent, but they really tried everything they could to see if there was a hydralazine-associated lupus syndrome in V-HeFT. And the summary of all of their work and deliberations is in the briefing document and there really isn't much there.

DR. YANCY: Jonathan, let me add two more observations. The first is that, as you point out, the risk of lupus drug-induced from hydralazine is, in fact, dose dependent and doses achieved in A-HeFT were significantly less. In fact, the maximum target dose was 225 of hydralazine and the dose achieved was less than that. Secondly, in

terms of reported cases of lupus in a trial, it was incredibly small, in the single digit range.

DR. SABOLINSKI: I would like to just call up a slide, please. On the left you see the adverse events for V-HeFT I and V-HeFT II in black patients. The third row shows arthralgias. We see that the I/H group had 65 and placebo had 61 and enalapril had 72. So, what I wanted to point out is that arthralgias were basically comparable in the isosorbide dinitrate/hydralazine group and the placebo group for both those trials.

DR. HIATT: Let me follow-up on what Jonathan was saying. I think this is actually really an excellent point. First of all, without dose ranging, we know you are at a relatively decent dose. My recollection is that 200 mg of hydralazine is kind sort of bordering on the unsafe zone and 400 is out there.

Secondly, you don't have the power to pick up these kind of rare events, and we have been burned before in this very area. So, this would speak I think--maybe this afternoon in talking

about postmarketing surveillance and methods--that may be if a more rigorous method had been applied to pick up what is potentially a significant risk in this population--I am not reassured by the absence of data. There is one case report in your safety section that talks about a patient who might have had it but that doesn't reassure me at all because you just don't have the power, and the confidence interval around that is huge. So, we are not excluding anything today.

DR. PACKER: I don't disagree. I should mention, of course, that concern about rare events and the wide confidence intervals around that is really a big concern when you are considering the approval of a drug based on a surrogate. But here you are considering a drug based on reduction in adverse outcomes, many of which are far more frequent than the adverse outcome that you are specifically concerned about.

DR. HIATT: I appreciate that fully. These are not symptomatic therapies. I totally appreciate that. But I think that relative to

other kinds of recent discussions, there should be the consideration of a more formal mechanism to pick up the signal. We are not going to have, you know, controlled data to do that going forward.

DR. SACKNER-BERNSTEIN: In terms of the comments, I realize that the issue I am raising has lots of caveats but I don't think the follow-up is purely what accounts for differences in arthralgia. It may just be the way people are reporting it but there is some imbalance in A-HeFT with arthralgias. It is about a four times higher rate, even though the numbers are very small, 1.5 versus 0.4 percent in A-HeFT. So, yes, arthralgias seem to be similar over time in the bigger study with longer follow-up, but the follow-up difference is 379 days versus 812 days. So, while it is twice the follow-up, I think it still falls into the category of concern that this is a risk that could increase over time. That is what the literature says, although even the figure of 200, which I have seen written, is one that seems to be based on someone's experience rather than actual data.

While the mean value was 169 mg below the threshold, now we are talking about an issue where potentially women would be taking the drug and, from the point of view of the risk that might be higher, from the point of view that their size in general tends to be smaller and, therefore, a toxic effect may be a higher risk, I realize that these kinds of rare things are not going to be the focus of development but there are serologic markers that are linked exquisitely closely to identifying a high risk population. And those serum markers weren't even looked at.

DR. PACKER: Two points. I want to emphasize that the biggest driving force for the difference in frequency of arthralgias in V-HeFT was actually not the duration of the trial, 60 percent versus the small number here. It is because in V-HeFT there was a checklist and patients were actually asked about each of those AEs at each visit and that is a materially different kind of procedure than asking patients to spontaneously report AEs.

The 60-odd percent of evoked arthralgias probably speaks to the fact that a lot of people have musculoskeletal pains every day and wouldn't necessarily report them as an AE unless someone says are you having it. So, it is a difference in terms of how the data were collected.

These are the actual specific data in A-HeFT. Let me emphasize that 40 patients in A-HeFT were women so, you know, there is a meaningful proportion of women here. These are all of the arthritis-related AEs in the safety population. I am not going to go through this. You can look at this and see if this raises a signal of provides reassurance.

DR. HIATT: Well, arthralgias are common and this is not unlike the statin-induced myopathy discussions. You know, there is a lot of discomfort out there in patients taking statins and there is very, very little documented rhabdom. So, once again, I don't think we can answer this question today. We just can't. As Jonathan has pointed out, there is no biomarker evidence of what

is going on. But I do think the committee should consider going forward on how to address these relatively rare but clearly drug-induced safety concerns. I don't think we can resolve that with the current data sets so we have to have some plan or discussion about how to go forward.

DR. NISSEN: Duly noted. It is now about 12:05 or 12:10. We are due for a lunch break. If people want to continue for a little bit longer, we can certainly do that and shorten our lunch break. I am trying to keep us on a time course today. Certainly, if the committee has questions they want answered right now we can do that. But my inclination is to break for lunch. We want to be back here at one o'clock sharp because we will undoubtedly have a very lively open public hearing. I am looking at the list of participants. So, I am going to start us at one o'clock exactly so if you want to hear what people have to say, you ought to be back here.

[Luncheon recess.]

A F T E R N O O N P R O C E E D I N G S

Open Public Hearing

DR. NISSEN: We are going to get started. We are actually a little bit late. I said I was going to start at one o'clock promptly and I broke my promise by being about three minutes late. We are about to enter the open public hearing portion of this discussion and I have a statement to read from the FDA:

Both the Food and Drug Administration and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the open public hearing session of the advisory committee meeting, FDA believes it is important to understand the context of an individual's presentation.

For this reason, the FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement to advise the committee of any financial relationship that you have with the sponsor, its products and, if known, its direct competitors. For example, this

financial information may include the sponsor's payment for your travel, lodging or other expenses in connection with your attendance at this meeting. Likewise, FDA encourages you, at the beginning of your statement, to advise the committee if you do not have any such financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

First up, we have Congresswoman, the Hon. Donna Christensen, who is Chair of the Congressional Black Caucus, who has requested time. If you would step up to the microphone, we would be pleased to hear from you.

DR. CHRISTENSEN: Good afternoon, members of the committee. I am here before you this afternoon as Chair of the Health Braintrust of the Congressional Caucus, and I want to say to you that today, ladies and gentlemen, you have before you an unprecedented opportunity to significantly reduce one of the major health disparities in the African American community and, in doing so, to begin a

process that will bring some degree of equity and justice to the American healthcare system.

Every day more than 200 African Americans die from premature causes. The leading causes of those deaths is heart disease which we suffer more disproportionately from than any other racial or ethnic group. Heart failure among African Americans is expected to increase from 725,000 to 900,000 in the next five years, and 50 percent of those patients survive less than five years after diagnosis.

Studies have suggested that deficiencies in nitric oxide play a role in congestive heart failure in African Americans, and the medication that we are discussing today, BiDil, widens blood vessels by increasing nitric oxide. Through the A-HeFT clinical trial with its 1100 African American men and women participants, it was found that the drug showed a remarkable 43 percent reduction in mortality, a 33 percent reduction in hospitalizations from heart failure, and an overall improvement in their quality of life.

It is important to note that the drugs which are included in BiDil are not new medicines. They have been standard treatment for heart disease and hypertension for decades. It is the specific combination of these drugs, known as BiDil, when used with other medicines which have proven by themselves ineffective in reducing mortality or improving quality of life, it is that combination that is before the panel today. So, I think we can assume that it is not the safety of the medication which is in question. In fact, the American Heart Association lauded BiDil as one of the top ten advances of 2004.

Neither would our concern be the A-HeFT itself because I think it could be considered a model trial for its methodology and the fact that, unlike some recent cases of medication already approved, the trial was stopped after 18 months because of higher mortality in the placebo group.

So, let me focus the rest of my remarks on the issue of the approval of the indication, the approval of BiDil for the treatment of congestive

heart failure in patients of one race, African Americans. When I spoke to the principal investigators last year I applauded them and the Association of Black Cardiologists and NitroMed for being willing to take what everyone knew would be a controversial step.

I didn't say this then but I also feel that to ignore the positive results in the few African Americans who were in the initial study would have been negligent. Today, because they took that risk having confidence in their product and ensuring that every care was taken to protect the interests of the cohorts, we are here asking for your approval for a drug that will save countless lives of African Americans, a drug we would not have had if they had ignored those findings.

So, why are we hesitating? This drug would not likely be approved for the larger population because it did not prove efficacious in whites who made up the vast majority of the first trial. Further, approving it for blacks today does

not prohibit further studies from being done in other groups. Neither does it pass any negative stigma on African Americans because it would be indicted specifically for us. We have been long stigmatized by any number of false assumptions and superficial traits where stigmatization is perpetuated even today and works to our disadvantage, denial of our rights and even death beginning, with the simple color of our skin. Would you deny a life now to us rather than do what the evidence shows can and should be done?

I have read some of the opponents' papers and I think many of the points of concern they raise are legitimate and offer some protections for future drug investigations and trials. We know that all of us, no matter what the color of our skin or race or ethnic origin, are 99 percent the same genetically. Approving BiDil as a drug for African Americans doesn't change that. Nowhere have I read in the study or subsequently heard that the choice of cohorts in the A-HeFT trial was based on genetics or any specific alleles. The

identification as black was self-described and that term, as we all agree, connotes not just less than one percent difference but it appears to override the far more genetic differences that exist among African Americans and black and self-described would also include all of the "social" forces and biologic feedback loops that Dr. Troy Duster admonishes us to understand.

The position at the CBC on the approval of BiDil is clear and unequivocal. It should be approved and indicted for use in African Americans. We are only cognizant of the many social, political and economic variance which define being an African American in the United States today. Addressing these in eliminating the disparities that exist in all aspects of our lives is our highest priority until those gaps are closed. Their continued existence despite our best efforts must not be used to deny treatment to those for whom treatment has been denied and deferred for 400 years. Today this panel is being asked to reverse that history.

Knowing that diseases are expressed

differently in different racial and ethnic groups, the challenge is not to avoid research but to act appropriately when this research is conducted and reported, and to commit to the continuous education of physicians and patients so that these drugs can be appropriately used. It is also critical that we continue the kind of research that was inherent in the promise of the decoding of the human genome whereby we move closer and closer to identifying targeted treatments and more precise measures than race for determining the effectiveness of a treatment.

Finally, it is our hope that the experience of A-HeFT and BiDil will encourage wider inclusion of minority patients and women in clinical trials, a position that the CBC has long encouraged and advocated. The results of A-HeFT could not be clearer in demonstrating that BiDil can save thousands of lives and reduce untold suffering for African American heart failure patients and their families.

I commend the FDA for encouraging the

inclusion of people of color in clinical trials. We encourage the FDA to do more. We also applaud their role in helping to design the A-HeFT trial to assess the safety and effectiveness of BiDil to treat heart failure in African American patients. I ask that you consider our perspective, the perspective of African American elected officials, in your review and decision on BiDil. Thank you.

DR. NISSEN: Thank you very much.

DR. CHRISTENSEN: And thank you for allowing me to come out of turn so that I can keep up with my schedule. Thank you very much.

DR. NISSEN: All right. We are going to move right along. Each of the speakers, by the way, has five minutes and I would like to ask you to stay within the five-minute allotted time. Our next speaker is Dr. Gary Puckrein, who is with the National Minority Health Foundation.

DR. PUCKREIN: Good afternoon. First of all a disclosure, the National Minority Health Foundation has received an unrestricted educational grant to undertake epidemiological research on

chronic heart failure patients from NitroMed.

The efficacy of BiDil has been researched over two decades, beginning with the Veterans Affairs vasodilator heart failure trials which were conducted in the Veterans Administration hospitals from 1980 through 1999, culminating with the African American heart trial which ended in 2004.

As evidenced by the A-HeFT results, approval of BiDil will have an immediate and positive impact on the health and quality of life of many patients with heart failure. Further, the lessons learned from the A-HeFT trial protocol will contribute to the experimental base required to advance progress towards personalized medicine and improve the quality of healthcare for all Americans.

In supporting the approval of BiDil based upon the A-HeFT trials, asserts no absolute or implied correlation between social, race, genetic type and the efficacy of BiDil. I support BiDil because it will extend the life of many Americans with heart failure. I support it because it will

improve the quality of life of these patients.

I understand for the purposes of the A-HeFT trial self-identified social race was used to define African American patient population. And analysis of the A-HeFT trial results demonstrates that a subset of the patient population responded favorably to BiDil. It is my understanding that the A-HeFT researchers do not assert that African American heart failure patients will be the only ones to benefit from BiDil, nor that the A-HeFT demonstrates that BiDil will not be effective in any other population groups that can be categorized by social race.

The results of the trials cannot be read to mean that it works only in African Americans of that it will not work in Caucasians or other racial groups. Further, it is my understanding that A-HeFT also demonstrates that adverse clinical effects were not presented in the patient population. Access to BiDil will reduce mortality rates and improve the quality of life for so many Americans, as well as lower the personal and

societal cost of heart failure. Conversely, lack of access to BiDil has the potential to create unavoidable human and physical resource demands on the healthcare delivery system and, most importantly, to unnecessarily compromise health status for thousands of American.

To attain this goal, BiDil must be part of the standard armamentarium of the treatment modalities available to physicians who treat patients with heart failure. We all recognize that the race and ethnic categories that we are currently using are not anthropological, meaning they are not scientifically based. Those categories described the sociocultural construct of our society. New science is compelling us all to delineate more precisely when and how these constructs can be evoked. For the purposes of the A-HeFT trial the compromise is made to those sociocultural constructs to identify a patient population who will have benefit from this new medicine.

Some geneticists and social scientists

denounce the combination as unscientific, but they cannot offer an immediate alternative to identify this population. Others suggest that we should wait until we have better categories. Our position is that we cannot allow people not to have their medications. It is important that this new medication be made available to all.

With that, I would like to offer my strongest endorsement for BiDil. Thank you.

DR. NISSEN: Thank you very much.

[Applause]

Our next speaker is Dr. Waine Kong, who is the CEO of the American Association of Black Cardiologists. Waine?

DR. KONG: Thank you and good afternoon. The Association of Black Cardiologists co-sponsored the trial and, as such, received funding for the staff support that we lent to the study, and also received funding for various projects that we have been undertaking since the study was completed.

As was indicated, my name is Waine Kong. I have been the CEO of the Association of Black

Cardiologists for the past 18 years. The ABC was founded in 1974, dedicated to the preposition that good health is the cornerstone of progress. The ABC is firmly resolved to make exemplary healthcare accessible and affordable to all in need; dedicated to lowering the higher rate of cardiovascular disease in minority populations; and committed to advocacy and diversity. We are guided by high ethics in all our transactions and strive for excellence in our training and skills.

The ABC recognizes the importance of partnerships in order to eliminate the cardiovascular healthcare disparities, made worse by lower socioeconomic status, access to and cost and quality of healthcare, and significant under-representation in clinical trials. The ABC-NitroMed partnership was initiated in 2000. The decision to partner with NitroMed was a source of much discussion at the ABC. However, after examining the available data and determining the potential benefits, we obtained consensus that we should move forward with this partnership.

We supported the study in the name of science. Our members felt that by direct trial participation, including principal investigator and subject recruitment, we would be able to confirm the data's validity for the medical community, particularly those caring for African American heart failure patients. We believed that the ABC would be doing a great disservice to the African American community by not obtaining the answer as to whether BiDil would reduce hospitalizations and mortality, as well as increase quality of life in our patients with heart failure.

Here is the background that made our partnership with NitroMed compelling. Heart failure affects approximately five million Americans and more than 750,000 of them are African American. Between the ages of 45 and 64, African Americans suffer from heart failure 2.5 times more than whites. Black patients are diagnosed with heart failure at a much younger age and die sooner than their white counterparts. When African Americans are diagnosed with heart failure, their

prognosis is generally poor, with 50 percent of them dying within five years.

The ABC contributed our expertise and our relationships with recruiting principal investigators. We helped to recruit patients. We organized and hosted community meetings to explain the study and to obtain the support of the community to make the trial successful. We are proud of what we have accomplished and what the study, in fact, taught us.

A-HeFT proved that significant mortality benefit can be gained in African American heart failure patients. We broke new ground in the pursuit of targeted treatments for specific populations and that will potentially save the lives of thousands.

In closing, the ABC is dedicated to assuring that children know their grandparents. This will not happen if we do not significantly impact on the mortality rate for cardiovascular diseases. That being said, the ABC strongly supports the approval of BiDil for the sake of our

children so that they can grow up knowing their grandparents. Thank you.

[Applause]

DR. NISSEN: Thank you, Dr. Kong. Our next speak is Debra Lee who was a patient in the A-HeFT trial.

MS. LEE: I would like for the committee to know that I am being reimbursed for my travel, but no one is paying me to speak today.

Hello. My name is Debra Lee. I am 48 years old and I have congestive heart failure. For those of you who don't know what this means, I would like to tell you my story. In 1999 I had a heart attack. There was blockage in my heart. A stent was inserted. In early 2003 I noticed changes in my health--coughing continuously; being visibly short of breath; walking short distances tired me out; waking up in the middle of the night; sleeping in a chair because I felt as if I would suffocate if I laid down. The doctors tested for various conditions.

In June, 2003 I failed a stress test,

showing the problem was my heart. In August of the same year it was confirmed. I had congestive heart failure. In September, the doctor offered me a chance to participate in a blind study, the African American heart failure trial. I quickly said yes.

More recently, in 2004, I was asked if I would participate in an extended version of that same study. Again I accepted. How do I feel now? I feel fabulous--no more shortness of breath; I am able to walk and exercise without resting; I can sleep in my bed at night; I am working more hours at the Indianapolis Museum of Art; I have more energy.

What I do contribute as the cause of this turnaround? It is my strong faith in God and a little pill called BiDil. I believe this pill is helping my heart to pump stronger. A normal heart pumps strong and steady. Patients with congestive heart failure, our hearts sometimes pump slower and irregular. In my opinion, this pill has changed so many things for me, given me a new lease on life. I have set new goals for myself. I know that

congestive heart failure and diabetes sometimes go hand-in-hand and one of my first goals is to control my diet and lose weight. I believe I have another 40 years or so to live my life to its fullest.

Just knowing that there is a pill that could help your heart to pump stronger--imagine how many thousands of patients this could help. Knowing that my mother had congestive heart failure, knowing she died because her heart was so weak it just gave out, I find myself wondering what would have happened if she could have had the benefit of this pill? Could it have helped her heart to pump stronger? Would she have lived a little longer? I feel I have been given an opportunity that my mother never had. I believe there are no mistakes; there are no accidents. I know God is working through this with me. I am sharing my story with you today because I feel BiDil can help other patients who live with the fear that your next breath will be your last.

I take 23 pills a day but my joy comes

from knowing that my medication is truly working its best to correct something that can't be fixed, my heart. If you would ask me, Debra Lee, how do you feel? I feel fabulous and I have to tell you I know I am blessed and God has blessed me with BiDil. Thank you for your time.

[Applause]

DR. NISSEN: Thank you. Our next speaker is Dr. Shomarka Keita. Dr. Keita is an anthropologist in the District of Columbia.

DR. KEITA: Good morning. This talk is entitled "BiDil: Patenting Blackness." I am concerned about the labeling of this drug as a black drug, something that I think will invariably happen and has already happened. I dedicate this talk, because it is June 16th, to Mark Bloch who was executed in 1944 on this day. He talked about the idyll of origins and he said in any study there lurks the danger of confusing ancestry with explanation. I also note that today commemorates the uprising in South Africa, where labeling has cost many lives of many people.

Next please. Principles--medications work at the levels of pathophysiology, clinical phenotypes and individuals and not on sociodemographic categories, groups or mystical identities. The African American group does not consist of uniform individuals or biologically the same, due to genealogical uniformity, or even environment insult. The race concept does not apply to modern humans. All aspects of adult biology are not inherited.

The formal correct race concept, subspecies concept, refers to particular levels of specific differentiation and evolutionary divergence. It differs from the incorrect colloquial use of race in sociopolitical and bureaucratic discourse. It does not apply to living humans.

Next. My position for recommendations is identifying the specific pathophysiology, clinical phenotype on which the components of BiDil work and their pathways using the components of BiDil to treat the susceptible clinical phenotype in any and

all individuals who have it irrespective of origin; using the clinical experience of those who have successfully used the components of BiDil in the past. They are not new drugs and some institutions have experience. Approving BiDil only if the labeling does not state that the drug is a race, black or African American, drug.

My position is against are labeling of BiDil in this manner. Approving BiDil in this fashion is scientifically unjustified in terms of population biology, leads to equating a social designation with a particular medicine as if otologically connected. It ignores the clinical experience of those who have had experience with these drugs and assumes that the developmental and later environmental causes of disease, namely social inequality, the biology and the poverty, will persist.

Next. Reasons against group labeling of BiDil--it has not been shown that the clinical phenotype that BiDil works on is exclusive to African Americans. I say presumably Afro North

Americans. We see to forget about Afro Latinos. Is it a connotation for all blacks? What about those from Morocco, Algeria, the Sahara, etc.? All adult biology is not inherited and group genealogically specific. The assumption when the words "race" or "genetic" are evoked to explain group differences--we must not ignore the developmental origins of disease or the later environmental and socioeconomic factors that influence clinical phenotype. There is no linkage of genes responsible for traits like variations in skin color with genes possibly connected to particular causes of heart failure.

Final slide. BiDil should not be approved if the labeling clearly states that it is not race--black, or African American--drug. This would be intellectually dishonest, amongst some other things. Thank you.

DR. NISSEN: Our next speaker is Dr. Jonathan Kahn, from Hamline University School of Law.

DR. KAHN: Thank you. I have no financial

interests one way or another in this.

First, I would like to commend the efforts of the many medical professionals involved in A-HeFT, and urge you to recommend approval of BiDil, but with one important caveat, approve BiDil for use in the general population without regard to race. There are several important reasons why BiDil should be approved for use in the general population without regard to race.

First, the data from A-HeFT support no claims that BiDil works differently or better in African Americans than in any other racial or ethnic group. This is because the trial enrolled only self-identified African Americans. There was no comparison population. There was, therefore, no scientific basis on which to claim race-specific efficacy for BiDil. Indeed, even the A-HeFT investigators concede that BiDil work in non-African Americans.

Second, arguments that data from V-HeFT I provide additional support for race-specific approval are unsupportable on several counts.

First, V-HeFT I enrolled only 180 African Americans. Second, an earlier version of this committee found, in 1997 when reviewing the V-HeFT data, that the statistics associated were too muddled to support a clear finding of efficacy one way or the other. A post hoc retrospective subgroup analysis of 20 year-old data cannot rectify this original statistical problem.

In this regard, it should be noted that several inaccurate claims regarding statistical difference between black and white mortality from heart failure have often been put forward to frame the BiDil application. Claims that blacks suffer mortality from heart failure at a rate twice that of whites are wrong. As I have shown in an article, published in Perspectives in Biology and Medicine, current data from the CDC indicate that black/white ratio of overall age-adjusted mortality from heart failure is approximately 1.8:1.0.

Claims that in the age range 45-64 blacks suffer mortality at a rate 2.5 times that of whites are accurate, but such claims leave out the fact

that that age group captures only about 6-7 percent of overall mortality from heart failure. In the age group 65 and above, where about 92 percent of mortality occurs, the ratio again approaches 1:1.

Third, under these circumstances approving BiDil as a race-specific drug would ratify the claim that race of the A-HeFT subjects was a relevant biological variable in assessing the efficacy of the drug, but the trial design produced no evidence to support such a claim.

I would like to emphasize that such a designation is fundamentally different from other labeling designations that suggest different dosages of varying degrees of efficacy based on probabilistic correlations with race of ethnicity. These latter designations serve merely as guides to individual physicians in calibrating drug dosage and administration.

A race-specific label for BiDil as a drug to treat heart failure in African Americans, however, would have far different consequences. First, rather than supplying information to guide

drug administration to particular individuals, it is a directive to doctors that this drug is for use only in African Americans. Any use of the drug in a non-African American would constitute off-label use.

Second, the argument that off-label use is a common practice and the drug will be readily available to allow doctors to prescribe it to non-African Americans merely indicates that BiDil, indeed, should be approved regardless of race.

Third, a race-specific indication for the drug would lead some doctors, quite reasonably, to think that this drug is not appropriate for many non-African American patients who, in fact, might benefit greatly from it.

Fourth, a race-specific indication might have a substantial impact on available insurance reimbursement for the use of BiDil, further restricting access to patients who might benefit from it.

Finally, most drugs on the market today were approved by the FDA based on trials conducted

almost exclusively in white patients, but these drugs are not designated as white drugs, and rightly so. Neither should BiDil be designated as a drug for African Americans. When approving drugs tested in white populations the proper assumption of the FDA was that the category "white people" did not differ in any meaningful way from the category "human being." The same assumption should apply to a drug tested in an African American population.

It is a simple idea that if the results of a trial conducted in a white population are good enough for everybody, then a drug tested in a black population should be good enough for everybody too. Thank you.

[Applause]

DR. NISSEN: Our next speaker is Dr. Charles Curry, from the International Society for Hypertension in Blacks.

DR. CURRY: Thank you very much. Good afternoon, everybody. I could almost say amen to what Dr. Kahn just said and sit right down. But I must first say that the International Society of

Hypertension in Blacks vigorously supports the approval of this drug. We feel that this is the most important advance in the care of black people that we have seen in my lifetime.

But leaving the issue of the International Society on Hypertension in Blacks and moving to Charles Curry, I must say, first of all, that I am a speaker for NitroMed and I was one of the investigators in the A-HeFT trial. I became an investigator after evaluating all the data and all the information that we have covered today, and concluded that, first of all, there could be no harm and probably would be very, very good for the black community.

I have been around longer than most cardiologists. I was chief of cardiology at Howard University for 25 years. In that period of time I have seen an enormous number of heart failure patients at Howard University Hospital and D.C. General Hospital and I am very much happy to have a drug like this come around. But during my years, I have also been able to see the development of

cardiology because it seems like everything started to happen when I happened. That is, I left Duke University and came to Howard in 1970 and studies started jumping.

But remember, think about this, the Framingham study is a very old, honored and respected study. There were almost no black people in the Framingham study but we have used that data and we have helped blacks and all races of people by using the data from the Framingham study that helped us understand the pathogenesis of cardiovascular disease. At one point they used to say that blacks did not respond to antihypertensive drugs. So, a hypertension detection and follow-up program was done. In that study there was a group of people referred to the community and a group studied at university hospitals and, lo and behold, blacks did just as well as whites. It was found basically that once you eliminated the socioeconomic barriers, in some cases when provided the medication blacks did well.

The 4-S trial is a trial that proved to

us--there were many doubters before the 4-S trial, that lowering cholesterol was a great thing to do. The Scandinavian--I forget what it is called--well, once that study was done the American cardiology community jumped on the statin drugs and they are really pulling them on now, and I think justifiably so. But would you restrict the results of the Scandinavian trial to Scandinavian people? I don't think so.

In my hand I hold a book which many of you have probably forgotten. It is "Clinical Practice Guideline: Heart Failure Evaluation and Care of Patients with Left Ventricular Systolic Dysfunction." It was published by the agency for Healthcare Policy and Research in 1994. On page 60, they talk about hydralazine/isosorbide dinitrate and they highlight the statement that hydralazine/isosorbide dinitrate is an appropriate alternative in patients with contraindications or intolerance to ACE inhibitors. This was not drawn up for black people. Many doctors, I will bet you in this room, have been using this combination for

many years, particularly since these guidelines came out. I made sure all my students had a copy of these guidelines, and it greatly helped reduce the mortality rate of heart failure patients, I think, at Howard University Hospital.

So, I think you can tell where I am going. I really think that it would be unfortunate if this drug was not approved, and it would be even more unfortunate if white patients and other ethnic groups were not allowed to have the advantages this drug seems to offer.

I want to close by saying a couple of things. One is that the A-HeFT trial I don't think should be expected to establish the mechanism by which the drug works. I don't think we really know for sure how the drug works, but what we do know is that there is a 43 percent reduction in mortality. Any doctor who treats sick patients cannot turn away from that.

The last thing I want to say is that over my many years I have seen dozens, or literally hundreds of trials comparing blacks and whites, but

almost invariably the groups are not well matched. The black population generally would have been sick longer with disease more well established and, yet they were the same age, with the same blood pressure and they said we have matched controls. So, the black population doesn't do as well. I think that is true in most studies you see when you compare African Americans with other populations. Thank you very much.

[Applause]

DR. NISSEN: Our next speaker is Basil Halliday, with the BDH Clinical Research Services.

MR. HALLIDAY: Good afternoon. My name is Basil Halliday and for the last ten years I have served as president and CEO of BDH Clinical Research Services. We specialize in doing ethnic-specific clinical trials. We participated in the A-HeFT clinical trial by providing a number of sites that had patients that participated in the clinical trial. I would like to offer a statement of support for BiDil, however, I would like to do that within the larger context of minority

participation in all clinical trials, not just the A-HeFT clinical trial.

Next slide, please. In so doing, I have entitled my talk "evidence that demands a verdict: BiDil as a case for increasing minority participation in clinical trials."

Next slide. As we look at the current picture, and I think in particular the industry view of the current picture, we find several things over and over. Frequently as we look at the package inserts we find that "no data is available" is frequently cited. As I have talked to project managers over the last 20 years--these are the people who are actually conducting the clinical trials, they say over and over we can't find them. Interestingly enough, we find them when we want to put them in prison.

Thirdly, as I have gone up the ladder of pharmaceutical companies and talked to the senior executives--and this is a direct quote, "we get our drugs approved anyway, so why bother?" The fact of the matter is when you put all this together we

have that minority participation in clinical trials averages less than five percent in trials supporting drug safety and efficacy, and it is time that we changed that.

Next slide, please. As we look at the same picture from the minority community perspective, frequently what we hear is "I don't want to be a guinea pig." That is often fueled by negative experiences with the healthcare system and an intense distrust of the industry, fueled by its poor image in the media.

Next slide, please. As we have talked to the FDA over the last 20 years, time and time again I have heard the statement "the FDA supports it; the FDA encourages it. We would like to see more of it." Yet, no mandate currently exists for increasing minority participation in the clinical trials process. Again, that is something that we must change.

Next slide. In understanding why we should all be concerned, most of us would agree that lack of minority participation in the clinical

trials process affects product development, standard of care, as well as outcomes. It also affects product approvals, even those based on non-U.S. data. At the very least, it ignores and rejects the question as to whether or not race and ethnicity do matter in pharmacological treatment. In short, it creates a vicious cycle.

Next slide, please. If you have poor trial recruitment and retention, it then forces you to overlook the differential impact of diseases by race, gender and ethnicity. Because clinical trials form the basis of modern medical practice, this overlook then forces a healthcare system that is unresponsive to the needs of the people that it is supposed to be serving. This perceived lack of responsiveness is then perceived as a lack of caring which then affects trust. You mix all this together and what you end up with is the stuff of health disparities. I submit to you that with approval of BiDil we can at least begin to break this cycle.

Next slide, please. When a physician

writes a prescription for a patient in his or her office, he basically has answered five questions for that patient: Is the product safe? Is the product effective? Is the dose correct? Is this the best therapy? But most importantly for me, not necessarily somebody else but for me, I submit to you that with approval of BiDil for the treatment of congestive heart failure in African American patients doctors will have a stronger confidence that the answer to all five of these questions is in fact yes.

Next slide, please. In the case of a drug where you have evidence that demands a verdict, I submit to you that BiDil demonstrates that race does, in fact, matter in pharmacological treatment; that, in fact, a representative sample was key to identifying superiority of BiDil in African American patients. BiDil will save African American lives and reduce health disparities.

In approving BiDil the FDA has a real opportunity to make available a drug that has been shown to benefit African Americans with CHF, a

population at high risk of disparate outcomes, including premature death. NitroMed's successful attempt to recruit African Americans in the A-HeFT clinical trial should serve as a model for considered efforts to recruit minorities usually unrepresented in clinical trials.

In supporting the approval of BiDil, I also ask that we take a look at the bigger picture and I ask that we all do a couple of things. From the minority community perspective, I ask that we become informed and consent to participate in the process not only as patients but also as investigators and also as advisory panel members. We need to become our own experts.

Number two, from the pharmaceutical industry perspective, increasing minority participation in all clinical trials is good business. It should not simply be a matter of lips service or tokenism.

Thirdly, from the FDA perspective, if we look at the data over the last 100 years it will point to us over and over again that there can only

be one answer to increasing minority participation in the clinical trials process, and that is to mandate it, mandate it, mandate it. Thank you for your time.

[Applause]

DR. NISSEN: Thank you. Our next speaker is Charles Rotimi, from the National Human Genome Center at Howard University.

DR. ROTIMI: Thank you very much. Listening to Debra Lee stand here and give her what I consider testimony to what BiDil has done for her, I just wanted to jump up and say "hallelujah". And also listening to Prof. Curry and the way he presented the case from a historical perspective, I absolutely agree with his position that it would be tragic not to approve this drug and it would be even more tragic just to approve it for African Americans.

With that, I want to inform the audience that we have a position paper that we have distributed to the members of the panel, and if you are interested we can provide it. It will probably

present a more coherent picture than I am about to display here.

But I would like for us to put this in the larger perspective in terms of health, not just the absence of heart failure. I think that if we do that we may appreciate some of the dilemma that we face in the way this process is going for BiDil approval for a specific drug for a particular ethnic group.

Next, please. I think if we look at the way the story is playing out in the media, it would be extremely naive on our part that this is not playing out in our own social notions of self-identity and group identity and what has put the minority populations in the United States at a huge disadvantage over the years. Nobody has stood up here and asked why do we have health disparity. Why do we have more heart failure in African Americans? Why do we have more hypertension? Why do we have more diabetes?

I came from Nigeria in 1981 to study in the United States and I have been here since there.

I married an African American. And it is extremely instructive for me to see that just for about all the conditions that I have studied as an epidemiologist African Americans are twice or higher risk. Why is that? Are they just selectively acquiring bad genes? I don't think so. There must be something in our social environment that drives health, that drives people towards poor health, and it is only by addressing that that we can truly reduce health disparity. That is the first point I want to make.

Also, this is the way it is playing out in the newspapers, that BiDil some day is going to lead us towards individualized medicine. How can we say that when, indeed, we are doing the very opposite? We are using group as a definition for the people that BiDil would be effective for.

Next slide, please. So, the first question that came to my mind when I heard the story about BiDil, I asked myself who is black? How do you begin to identify that? Even the concept of African American? At Howard University

in interacting with my colleagues I said that for me, the only way I can truly define that is the descendants of the ugly history of slave trade, of the middle passage; I don't have any genetic way that I can consistently draw a cycle of how all African Americans are in that cycle. Therefore, the label that we use is sociocultural and is derived from the ugly history of slave trade. If we are not conscious about that in this process we are going to exacerbate that whole social phenomenon, that group identity is confused with ancestry and that African Americans have multiple ancestry, and we must consider that when we are talking about biology.

Next, please. We are going from African American and we are now using the concept of black. Black is a big experience, a global experience, more than the United States. So, if my mother comes to the United States or a cardiologist in Nigeria wants to use BiDil they would define it as it is okay, she is black; you can put her on it. Has that been tested within this structure? No.

The whole concept of group labeling misleads us in biomedical research. For example, that data that is shown there shows that in Nigeria hypertension is about 7 percent, whereas in the Caribbean it is about 21 percent and among African Americans it is about 34 percent. These are men and women who are 25-34 years of age. So, clearly, the environment you find yourself drives the phenomenon.

Next slide, please. This is recently reported meta-analysis by Richard Cooper and co. When you hear a discussion of heart failure, hypertension and all that business in the United States you begin to think that there is something unique about the African American experience. What you see here is that the African American is right in the middle of the distribution of hypertension in the global context. So, there is no uniqueness there. You have black populations in different parts of this distribution.

Next, please. It is the same story with ACE inhibitors. Clearly, if somebody was interpreting the ACE findings from an unbiased

position, who has not been polluted by the social context within which we are trained, would say that most people respond to ACE inhibitors.

Next slide, please. These next two slides really indicate what our positions are. We advocate that if BiDil is going to be approved it should be approved for everybody, with a clear indication that it does not replace all therapy and that it may work, indeed, for different populations because it wasn't tested in this study.

BiDil, again, should not be labeled as a black drug. Okay? The point is that Prof. Khan made earlier, that it looks like we have come full cycle in biomedical research. We used to do only white males and that drug applied to everybody. Now why are we changing the paradigm for the black population? I think it is a critical question we must ask.

Next slide, please. So, promoting our health is what we should concentrate on, not just absence of disease. What good is a drug that reduces mortality from heart failure by 43 percent

if, because of patent, the drug becomes inaccessible because of increasing cost? Would it result in increase in hypertension and associated complications, including heart failure, due to increase in psychosocial stressors? Is biomedical research down the wrong path by suggesting without proper scientific justification that the so-called racial categories are biological? Thank you very much.

[Applause]

DR. NISSEN: Thank you. Our next speaker is Charmaine Royal, also from the National Human Genome Center at Howard University.

DR. ROYAL: I have no financial interests to declare. I have two slides, and I entitled my comments the "Illusion of Inclusion." I am hoping in my five minutes to challenge us all to look beyond the surface of inclusion. When the A-HeFT the trial ended last year and the debate really started, I began to think that the discussions about the drug and whether the drug should be approved for African Americans. In my opinion, it

centered around this whole issue of inclusion. The development of the study, the planning of the A-HeFT trial seemed to address the issue of inclusion of African Americans and other minorities in clinical trials. It also seemed to address the issue of health disparity.

Then, when the study ended and the general community celebrated, the African American community, many people in the African American community were thinking finally we have our drug; something for us. Somebody is paying attention to us, the whole issue of inclusion being part of the process, finally being part of the process.

Then, the advocacy groups were saying approve BiDil. It works for black people; let's approve this drug. Then I imagine for the FDA is also the issue of inclusion--we heard it already this morning about inclusion and what that means and BiDil and the A-HeFT trial being responsive to that.

I want us to look a little below the surface. On the surface, of course, inclusion is a

great thing and is something that should be applauded, and the organizers and investigators of the A-HeFT trial certainly need to be applauded for planning and implementing that study. But one of the first things I am going to talk about in terms of inclusion and what I see as part of this illusion, particularly if BiDil is approved as a drug to treat heart failure in African American, is the issue of truth-telling. We have heard a lot of discussion about what we know about populations and that genetic variation is overlapping, is continuous. We don't have discrete boundaries between groups and labeling the drug as a drug to treat African Americans implies that, that there is something about this group of people that we call African Americans that makes them different from another group of people and so we need to focus on these people.

I think the FDA evaluation process, and other agencies too, need to hold investigators accountable for conveying the truth about what we know about populations. Certainly including

populations is good, but what do we with the results that we get? That is where we are with BiDil. The study has shown efficacy in this population, but do we market it as the drug only for African Americans?

Jon Kahn and others talked about the departure or the seeming departure from what we have done in the past with approving drugs, and the whole issue of studies being conducted in one group and being applied to others. Over time, the ethical theory of utilitarianism is what has guided much of what we do--the greatest good for the greatest number of people. Certainly, limiting BiDil to African Americans will not allow us to accomplish that. And, why should BiDil be a departure from the way we have done this before in terms of benefiting all people for whom the drug could work?

The whole issue of access, how much will BiDil cost as opposed to the component of BiDil which other physicians have been using for a long time? In my conversations with some physicians the

cost of BiDil probably will be three to four times the cost of the components that they have been using over time. Will the African Americans, the target group, be able to afford this drug?

Then, what about other people from other groups for whom this drug may work? Are we going to deny those people the benefit that would heard Debra Lee talk about, the benefit of this drug?

Last but by no means least, how are we going to implement this in the clinical setting? Who is African American? How are we going to identify African Americans? How is the decision going to be made about who is black? Are we going to allow people to self-identify? Is the physician going to be the one that says you are black? The question about identity is one that is critical here. Are there going to be criteria, national standardized criteria for how people identify individuals for the treatment of BiDil?

I really hope that the FDA and that all of us here will really think critically about what will happen, and the ramifications of approving

this drug for the treatment of heart failure only in African Americans, because in think about inclusion, in my mind, will be just an illusion. Thank you.

[Applause]

DR. NISSEN: Thank you. Our next speaker is Olivia Carter-Pokras, and that is also listed as Kendrick Gwynn so I suspect there will be two people involved.

MR. KENDRICK: Good afternoon. I am here representing my mentor, Dr. Olivia Carter-Pokras and we have no financial obligations.

I would first like to thank the members of the advisory committee for allowing me the opportunity to provide comments on BiDil. I am here today to present the results of a research study that I did that is currently under review and is relevant to your deliberations.

As a student at the University of Maryland School of Medicine, I study how race and ethnicity is used for clinical trials, marketing and dosage recommendations for cardiovascular drugs prescribed

to African Americans.

During her work for the review of federal standards for racial and ethnic data, Dr. Carter-Pokras learned that the package insert for an antihypertensive drug that recommended twice the dosage for blacks than for whites. The executive director of Project Race, a multiracial advocacy group expressed concern about how physicians will care for multiracial children.

When Dr. Carter-Pokras brought this to my attention we started a three-part research project. First we examined the "Physician Desk Reference" for any references to race and dosage recommendations and adverse events. Physicians are advised to follow the recommendations contained within these materials in treating their patients but are not restricted from deviating from these guidelines.

Several core cases have upheld the PDR as the legal standard of care. Second, we reviewed public comments submitted to the FDA regarding the 2003 draft guidance for industry on the collection

of racial and ethnic data in clinical trials. Third, we conducted in-depth interviews to obtain views on the definition of race and ethnicity in the pharmaceutical industry. We obtained it from the University of Maryland School of Medicine's institutional review board. Out of 135 cardiovascular agents, only one ACE inhibitor had dosage recommendations that varied by race, suggesting 2 mg for blacks and 1 mg for non-blacks. In addition, we found mention of a higher incidence of angioedema among all ACE inhibitors for all blacks.

We interviewed 11 informants who remain anonymous but who represented the FDA, the National Pharmaceutical Council, Association of Black Cardiologists and a broad range of research fields. We asked them seven questions. How do you define race and/or ethnicity? How is race and ethnicity used by pharmaceutical companies? What are the reasons for these uses? Do you think that the knowledge from the Human Genome Project will change companies' use of race and ethnicity? How should

race and ethnicity be used in recruitment for pharmaceutical clinical trials? What are your views regarding clinical trials that include only one racial group? What are your views regarding marketing specific drugs for African Americans? What do you think the role of the pharmaceutical industry is, if any, in eliminating health disparities?

Pharmaceutical companies also addressed their views in the public comments to the FDA draft guidelines. Douglas Leasepoint[?] of Abbott Laboratories stated that the designation of race/ethnicity categories as sociocultural rather than anthropologic, while politically correct, weakens the utility of genetically influenced differences between populations.

Our informants agreed on the definition of race as a sociopolitical construct. They were inconsistent on whether the Human Genome Project would impact how drug companies use race. One informant noted that clinical trials with only one racial group should be the exception and not the

rule. They urged that clinical trials include members from all races and ethnicities. Informants also agreed that certain racial groups should only be targeted for marketing of drugs if there is scientific evidence to support it. Finally, key informants felt that pharmaceutical companies do have a role in eliminating health disparities.

In summary, we found that the use of race and ethnicity for cardiovascular agents prescribed to African Americans is inconsistent with the social view of race. Although the 11 in-depth interviews we conducted is considered a reasonable number for qualitative research, we cannot say that the views expressed by our participants are a full representation of the diversity of views in medical research and pharmaceutical communities. We recommend that a pilot study be replicated to get a broader view of these issues. Thank you.

[Applause]

DR. NISSEN: Thank you. Our next speaker is Lucille Norville Perez. Dr. Perez is the NAACP national health director.

DR. PEREZ: Thank you. Ladies and gentlemen of the Cardiovascular and Renal Drugs Advisory Committee, thank you for the privilege of continuing and participating in this important process, and I ditto just about everything that I have heard thus far so my remarks will be short. I have been constantly crossing out but not diminishing their importance.

As has been said, I am Dr. Lucy Perez and I am a past president of the National Medical Association and the organization for which I am also speaking this afternoon on behalf of over 30,000 African American physicians and hundreds of thousands of African Americans burdened with heart disease all over the nation, I would like to thank you for having open-mindedness in considering this promising therapy.

Many previous speakers spoke of the importance of this today. This is my mother's birthday. My mother died 23 years ago and she died because of complications of diabetes and heart disease. So, this is personal.

The NDA in question, 20-727, quotes official FDA correspondence to NitroMed in 2001, wherein it stated the following: Given the subset finding and the overall trend toward a survival effect in V-HeFT I, we believe a single, clearly positive study in a black congestive heart failure population would be a basis for approval of BiDil for the treatment of heart failure in blacks.

Per the FDA suggestion, such a trial has been conducted and the results are now a matter of public record. We have spent all day discussing it. And, I think it is important to point out that this model study moved us closer to what we should have been doing with all drugs a long time ago.

As you have already heard, patients receiving BiDil in addition to current standard therapies compared to patients receiving current therapies and a placebo experienced a 43 percent--a 43 percent reduction in mortality; 39 percent reduction in first hospitalization for heart failure; and, as we heard from Debra Lee, a much improved quality of life.

Given that the p values are statistically significant and they were established without question, and given that this trial meets the standard set by the FDA of a single, clearly positive study in a black congestive heart failure population, this study's results should not be obscured by invalid ethical concerns of perceived political objections.

Given that you are convinced that given the disproportionate impact of cardiovascular disease on African Americans, anything short of approval of BiDil for use in this population cannot be justified and would be tantamount to the FDA disavowing its written and totally sound commitment in 2001. The National Medical Association, therefore, urges this committee to recommend to the FDA that BiDil be approved. We join several other organizations in this request, including the International Society of Hypertension in Blacks, ABC, the NAACP, the Alliance of Minority Medical Associations, the National Minority Health Foundation, several others are also aligned with us

in supporting the approval of BiDil.

African Americans continue to die, as my mother and my father, from heart disease at the alarming rate of 78,000 a year. This number could be significantly reduced if BiDil is brought to market as soon as possible. I intend, as you heard from Dr. Kahn, to do the electric slide at my great-great-granddaughter's wedding. My children will know their grandmother.

[Applause]

DR. NISSEN: Thank you, Dr. Perez. We have one final speaker and the last word goes to Donna Wells, who is a patient from the A-HeFT trial.

MS. WELLS: Good afternoon, ladies and gentlemen. I am Dianna Wells, a participant of the A-HeFT study. I identify my progress on BiDil by the measurements of a football game. I have made several touchdowns and field goals. I became short-winded when working around the house or simply walking down the street. Today I can walk up and down stairs and I also do normal housework

with ease. This is a major touchdown. At night I had to sleep on three to four pillows propped behind my back and neck and when I would get comfortable I would end up waking up 30 minutes later gasping for air. Now I can sleep on one pillow. This is a first down. I experienced excessive bloating. My clothes fitted tight. I was forced to wear a larger shoe size. I went to the emergency room where the doctors were even planning on cutting my jeans off because they were so tight from all the swelling. After being misdiagnosed with bronchitis and the flu, I was diagnosed with heart failure in the emergency room. As an inpatient in the hospital, the doctor was considering a heart transplant and I wasn't ready to accept this option.

I was unable to work, and after a few months on BiDil I have returned to work. I also volunteer at the library and I work at a food bank. This was a field goal that turned into a touchdown. Every game has a referee with colorful flags. My side effects were minimum. The dry skin is the

yellow flag. The headaches from the BiDil is the red flag. I remember the quote, "I am the master of my faith and captain of my soul." I use this phrase to overcome the fear of being a heart failure patient.

I studied and I also self-educated myself on my illness. I contacted the New York Chapter of the American Heart Association. They informed me of the study called A-HeFT. I later contacted my doctor, who is Dr. Martin Burke. Having the option to use this drug is better than waiting on a hot transplant which will usually cost over $55,000. Taking this pill for the rest of my life will cost less. It will mean a continued, productive life. I can't imagine my life without BiDil. Please take this into consideration when you are making your decision.

I would like to thank the BiDil sponsors and also thank you for this time to speak. As a minor note, I am only being reimbursed for my travel expenses. Thank you.

[Applause]

DR. NISSEN: I want to thank all of the speakers. It is actually very helpful to the committee to hear different perspectives and understand where particularly patients are coming from and the ethical issues raised are certainly very important and have to be considered by the committee.

Now I would like to move us, if it is the pleasure of the committee, unless there are burning questions for the sponsor, into the question session, and so just everybody understands the ground rules, this session is for the committee. So, we are not going to entertain any comment from the sponsor or from the audience unless we are asked to do so by the committee. Before I do that I want to make sure that the committee has all the questions that you want answered from the sponsor. So, before I close out that portion of the meeting, are there any burning questions for the sponsor, clarifications that we need before we move forward? Please go ahead, if you have them. John?

Questions to the Sponsor

DR. TEERLINK: I do actually have a question for the sponsor and it is in regard to the compound nature of the drug. On the one hand, we are given isosorbide dinitrate as a venodilator and the hydralazine as an arterial vasodilator, which is the one kind of way to spin it, in which case one would think that those would be reasonably appropriate to any kind of group I think. Then, the other kind of way to look at is that you have complementary effects where you have the nitrate donor aspect of isosorbide dinitrate that is then preserved by this concept of hydralazine preserving the nitric oxide aspect of isosorbide dinitrate.

The different roles of this agent, and understanding the different roles of this agent, could potentially have impacts on how broad of an indication you believe it is appropriate to give. As has been suggested I believe in the sponsor's prespecified, if you believe that within African Americans there is an enriched group of people who have this nitrate intolerance, then if, in fact, you believe that this works by maintaining the

nitrate balance pathway, then it would be more likely to work better in that population. If, in fact, you think it is a balanced vasodilator and that this is yet another study of balanced vasodilators, then there is less reason to believe that it would have an ethnic specificity.

I would be interested in hearing what the sponsor's view is of this in terms of how we should be looking at the agent from this perspective.

DR. WORCEL: Manuel Worcel, from NitroMed. I agree with you that from what we know from the literature the two components of the BiDil combination behave in the way you described. You get the vasodilating properties of isosorbide dinitrate and the complementary effects of hydralazine which is an arterial dilator. It has the potential to actually protect the metabolization of nitric oxide and then really complement nitric oxide availability in cases of nitric oxide dysfunction.

This being said, you asked us what we believe as sponsors--

DR. TEERLINK: And what the clinical data is to support that.

DR. WORCEL: Essentially, this is what I was discussing, that we believe in the data we submitted to the advisory committee today and what we consider supports the indication is the total database on the data and reports presented to the committee. We believe that the mechanism is different from just being two vasodilators. But from there to interpret the clinical data is a way that I wouldn't go.

DR. NISSEN: For evidence, is that what you--

DR. TEERLINK: No, I am interested. There was a very nice article done in 1980 by Dr. Cohn who looked at 3-month administration of oral nitrates. Actually, the title is "Sustained Hemodynamic Effects without Tolerance During Long-Term Isosorbide Dinitrate Treatment of Chronic Left Ventricular Failure." So, I am interested in seeing that article that suggested that there is not tolerance to oral nitrates in heart failure

patients. We have all assumed that there is tolerance in heart failure patients to oral therapies. I have seen a lot of articles in regards to intravenous administration of nitrates. Then, in addition to that, I would love to hear kind of what the evidence is that hydralazine reverses that tolerance effect, or is this really just the two balanced vasodilators?

DR. PACKER: John, I can try to address part of your question. We actually did, a long time ago, a study looking at the development of tolerance to isosorbide dinitrate comparing twice daily dosing, three times daily dosing and four times daily dosing in patients with heart failure, using invasive hemodynamic measurements. We actually published this as an abstract. I am embarrassed to say that we never translated it into a full manuscript.

Having said that, the doses we used were 40 mg 4 times daily, 40 mg 3 times daily, 40 mg twice daily, and we showed that the only regimen where there was no attenuation as 40 mg every 12

hours. When I say twice daily it was every 12 hours. And 40 mg every 8 hours produced an attenuation; 4 mg every 6 hours produced an marked attenuation. And that is probably the best data that exists, using different regimens of nitrates, showing that nitrate tolerance occurs with frequent administration in people with heart failure.

DR. TEERLINK: Dr. Cohn's data with 40 mg q.i.d. at 3 months still decreased pulmonary capillary wedge pressure, decreased mean arterial pressure, decreased systemic vascular systems.

DR. COHN: Well, I think this is a very interesting discussion obviously, John, but not pertinent really to today's regulatory issues. But let me just address a couple of your questions. In the data you refer to, we did give a holiday overnight so that we were studying the effect in the morning so there was a window of about 8 hours before after they had taken their last dose when they were still responsive to the next dose. So, it is probably not too dissimilar from the kind of data that Milton provided.

But you asked several different questions here. One important observation is that maintaining hemodynamic effect does not necessarily imply that you are maintaining efficacy based on the long-term activity of the drug. I think the best example really is V-HeFT I with prazosin, which is also an arterial and venodilator and produces the same hemodynamic effects really as isosorbide dinitrate and hydralazine, which is why we included that in the V-HeFT I trial. The blood pressure effects of prazosin persisted throughout the entire follow-up period, suggesting that that drug maintained its hemodynamic effects. Isosorbide dinitrate and hydralazine did not have a sustained blood pressure lowering effect in V-HeFT I. The blood pressure did not change. Prazosin had no influence on mortality or on left ventricular remodeling despite its clear-cut arterial and venodilator effects. Nitrate and hydralazine inhibited remodeling. The ejection fraction went up and stayed up and, as you have seen, had a rather profound effect on survival.

So, I don't think one can equate hemodynamic effects with long-term efficacy. Not all drugs which dilate will have a favorable effect on outcome. And drugs which have a favorable effect on outcome, like beta-blockers, do not dilate. So, I believe we are looking at drug specific effects and at the moment what we know is that the combination of isosorbide dinitrate and hydralazine, as in the BiDil preparation, is effective on both remodeling of the left ventricle, and we haven't been able to present that data today, and certainly on all the outcome measurements in heart failure. I don't think we can go back and look at the components. I think Bob Temple suggested that earlier. We remain, obviously, not certain about that.

DR. TEERLINK: Yes, I was just looking for some guidance to help us decide to whom to give it if we did decide to approve it.

DR. NISSEN: Jonathan, I think you wanted to ask some questions of the sponsor.

DR. SACKNER-BERNSTEIN: Just one right

now. I am just wondering if you can provide information on how many of the patients in each of the treatment groups had been treated with vasodilators, specifically hydralazine and nitrates, at some point in time in the past prior to study enrollment as opposed to the baseline medications.

DR. TAYLOR: That was a specific exclusion criterion. So, if the patient required hydralazine or nitrates ongoing any time in the past they were excluded from the trial. I would ask if we have data on the past history to reflect usage in the past. We don't.

DR. NISSEN: It sounds like it was an exclusion.

DR. TAYLOR: Yes.

DR. NISSEN: Other questions, factual questions for the sponsor? I want to make sure you get your questions answered.

DR. PACKER: Steven, with respect to the general policy of the committee, do you want us to address question two?

DR. NISSEN: No.

DR. PACKER: Okay.

DR. NISSEN: Bob?

DR. TEMPLE: As I said a number of times to no response, it seems entirely relevant to the question being raised, and it was raised by some of the people who spoke about a possible broader indication in the absence of any specific evidence, to go back and look at V-HeFT I for the results in the white self-designated population. My impression of it is that if anybody found results like that in their Phase II studies they would abandon a drug, but maybe I am exaggerating how strongly negative they are. So, maybe the committee doesn't feel it needs any more information on that, in which case tell me and I will withdraw the question. But if that is relevant, all Milton presented was the results in the black subset which, arguably, supports A-HeFT but does not go to the question of the lack of evidence for, say, carrying out a study in a white population, at least at that time.

DR. NISSEN: I guess the reason I hadn't asked for it is that there wasn't a statistically robust effect in the entire group and since the black group had the greatest benefit one can infer that there is not much effect in whites.

DR. TEMPLE: Well, not much but it may turn out to matter given the comments that people have made how strong that is.

DR. NISSEN: If you have it and we can do it quickly--

DR. PACKER: I am just going to summarize very quickly that what you see is the point estimate in white patients. Bob's question about whether a pharmaceutical company would develop the drug is sort of an interesting thought experiment. Let me just put out two pieces of information. In white patients in V-HeFT I there is a 12 percent point estimate reduction risk, wide confidence intervals. What you need to put alongside of that is that in V-HeFT II in white patients there is a 39 percent greater risk compared to enalapril. So, just put those two together in terms of the

thinking process.

Committee Discussion and Questions

DR. NISSEN: Other factual questions from the committee or may we get into the questions? I am concerned because we do need some time to discuss the questions but I don't want to cut anybody off. Let's get into the questions, into the meat of things. Again, I am going to enforce keeping the discussion to the committee so we have adequate time to discuss this amongst ourselves. But if you want to call upon the sponsor, you can but I am not going to encourage it.

The committee is asked to opine on whether V-HeFT I, V-HeFT II and A-HeFT adequately support a claim that BiDil, hydralazine plus isosorbide dinitrate, improves outcome in patients with heart failure. The advisory committee previously reviewed V-HeFT I and II as a possible basis for use of BiDil in the treatment of heart failure.

Claims based on A-HeFT: The primary endpoint was a composite of all-cause mortality, hospitalizations for heart failure and response to

the Minnesota Living with Heart Failure questionnaire. By the sponsor's and the statistical reviewer's intent-to-treat analyses, BiDil was associated with an improved composite risk score, p value equals 0.021 by the reviewer--that is the FDA reviewer. However, the sponsor's prespecified per-protocol analysis is not significant, p equals 0.46.

1.1.1, why are these results so discrepant? Anybody want to take that? Tom, you have drilled down on that.

DR. FLEMING: The per-protocol analysis that was done, as we are noting here in 1.1.2 and as the sponsor confirmed, excluded 60 percent of the ITT population. As you would expect and was clearly confirmed by Lloyd Fisher, these exclusions were not at all at random. The potential for major bias in directions that aren't always easy to predict can be anticipated when you have such substantial exclusions. So, it really renders any such analysis uninterpretable and the only criticism I would make is why the sponsor even

proposed this in the first place. But to defend the sponsor, fortunately, it wasn't one of their primary analyses; it was one of their sensitivity analyses.

So, the bottom line is I think this issue is not a key issue. I think the per-protocol analysis, as you would expect and certainly in this case, is essentially uninterpretable.

DR. NISSEN: Let me also just chime in and say that I am really not terribly interested in the per-protocol analysis either. The intent-to-treat analysis is the valid analysis, the one that should be used and I focus my attention on it. I think that the reasons that you stated also answer 1.1.2 and, unless anybody has any additional comments, is that adequate from your point of view? You know, I think we recognize that sensitivity analyses are sometimes useful but this one is particularly colored by the 60 percent exclusion which gives it very little power.

Let's move on to 1.2. Subjects enrolled prior to the second interim analysis, when the

sample size as re-estimated, comprised 30 percent of the total patients and 24 percent of events, and they showed a 7 percent nominal lower risk of death on BiDil. Subjects enrolled after the second interim analysis had a nominal 62 percent lower risk of death on BiDil. How troubling is that difference? How comforted are you by, 1.2.1., more continuous analyses of mortality by time in study?

That is really somewhat of a statistical question but, Tom, I think you might be a good person to answer that.

DR. FLEMING: 1.1. was easier. Can I answer that one again?

[Laughter]

DR. NISSEN: No.

DR. FLEMING: Well, this one is a tougher issue and one that I do want to spend a little time commenting on and give a little bit of background in the response. I am pleased Dave DeMets is here. In fact, I am very pleased that the DMC was able to be guided by his experience leadership to address these ongoing complicated issues that often occur

in trials, as did occur in this case.

There are a few places in the FDA reviews that I am focusing on here in my comments. The first of these places is page 26 in the medical reviewer's summary where the medical reviewer provided, in essence, kind of a description of the chronology of what happened during the course of the monitoring of the trial and Dave was in essence expanding on some of this.

Essentially, while the trial had originally been planned to have three looks, the monitoring committee essentially appears to have had approximately four or five key meetings during the course of the study. In March of '03 they initially met and then five months later they had their first look. I would have argued at that point the DMC should have been unblinded. Then, in the March of '03 analysis meeting they had their second look. According to what is identified here, in this meeting in March of '03 the committee unexpectedly unblinded itself for a second look and it was concluded that the treatment difference was

small but favorable to BiDil.

Now, we understand this is when there was an adjustment in sample size and, in essence, the process that was used--and it gets back to Bill's questions earlier on today--has implications in how we interpret the data. It is very appropriate and at this point, with the refined methodologies that exist, totally straightforward to address the fact that if you target a certain treatment effect size for a given power at a given false-positive error rate that defines the sample size. But it is based on an understanding of what the variability is in your estimates. If that variability estimate is not correct you can use the data to address that variability inaccuracy and adjust the sample size, and that is not problematic at all.

In a time-to-event trial it would be synonymous to saying what is fixed in the trial is the number of events. So, if you were targeting a 50 percent reduction with 80 percent power with an 0.25 false-positive error rate, that would set the specific number of events, which I think in this

setting here would have been about 8. But what isn't known is the exact sample size and that is what you can address as you are monitoring the data.

What becomes problematic is if you also adjust the sample size based on the emerging event treatment differences that you are seeing. So, it then becomes, in essence, in some level data driven. One has to factor that in, and I will come back to that in a minute.

The committee then met again a year later and then, seeing differences emerging in survival, recognized that it could readily be a case where early termination might be ethically and scientifically indicated because of those emerging survival differences, and they implemented a Lan and DeMets implementation of O'Brien-Fleming. That is post hoc. Would that concern me? Yes, it would except the fact that it is pretty ubiquitous, i.e., that would be like saying we forgot to say what test statistic we are going to use for time-to-event data; oh, let's make it log-rank.

Well, since log-rank is the norm it would bother me relatively little if somebody forgot to specify the method and then specified the one that is almost always used. So, the post hoc implementation of Lan and DeMets implementation of O'Brien-Fleming doesn't bother me greatly but I will come back to that.

The committee then met three months later and then recommended termination. That was unexpected but with the flexibility of Lan and DeMets that is, in fact built in. That is acceptable. So, in essence, to a great extent what was done under the data monitoring committee's guidance made sense, and it made sense for patient protection as it made sense for ethics. What is problematic is that, from a perspective of interpreting statistical strength of evidence, there is a considerable risk is that there is a complication in that interpretation when we are using the data part way through the course to essentially refine the hypothesis. You are not refining the hypothesis when you change the number

of participants to get a targeted number of events, but you are changing it when you change the number of events.

Essentially, what has emerged are what are called adaptive methods, methods that have been put forward to try to address this issue. There are a number of different ways to do it and I am going to try not to take a lot of time here but at least I want to describe some of the core elements of the thinking so that people have at least a little intuition.

In this trial initially there wee 300 patients and they were supposed to get 300 more. If you look at the statistical review and look at the FDA guidance document on the statistical review, on page 15 you can see the reviewer's summary of the data that exist at that time of that second look, which is when the relative risks were looked at and sample sizes were changed and the data that occurred after that point. This is page 15 in the statistical review.

Now, in a nutshell here, what is

happening--so, as you are trying to get 600 people you got 300. You look at the data. Suppose the treatment effect size is more modest than you had hoped so that you would have liked to have had a bigger trial so let's just add another 600 so we now make it 1200. That is not quite what they did; they added 500 but I am just going to make it even numbers here. So, you add another 600. In essence, the effect size that you see in the first part is critical here. The effect size in the first part, you notice the relative risk is 0.93. What we are really estimating is the log-hazard so the negative log of 0.93 is 0.07. Okay? In the second half it is 0.38. So, that effect size, when you take the negative log is 0.97. Essentially, the effect size, as statisticians are estimating, is 15 times larger in the second half than the first half.

What we know from multiple testing is that you can't enroll a sample size and test, and if it is not significant add 30 and test, add 30 and test, add 30 and test. In fact, if you do that

eventually, even when there is no effect with probability 1, you will eventually achieve an 0.05. So, we know we have to adjust so adjustments can, in fact, be achieved. One way to adjust is to recognize that if you enroll the first 300 you are locked in. Your intention was that was to be the first half of the information so you are locked in to keep that, the first half of what you use. You have to give that half the weight. So, if you add another 600 so you have 900 in the second half normally we would give three-quarters the weight to the second half. You have to still give that half the weight. Instead of giving only a quarter of weight to the first half, you are still locked in to giving it half the weight.

That is overly simplistic but it is essentially the principle. That is how you can look at the data and make refinements but you are not able to down-weight or de-emphasize what you have already seen. So, in simplicity, if you estimated an effect of 1 in the first half and 15 in the second you are inclined, because it is

three-quarters of the people that you gave 15, to estimate the effect as 11.5. No, the answer is 8. You have to give half the weight to the first half, half the weight to the second half.

So, it is more complicated here because, in essence, in this setting it is numbers of events, it is not numbers of people. Essentially, in the first half there were 36 events. In the second half the number of events is actually larger than what they saw because there was early termination and there were 50 events. The potential events, by my calculation, were 68.5.

So, without taking you through what was about a 3-hour calculation here a few nights ago, essentially one is in a position that the weight that you are giving to the second half is substantially less because of the fact that you used a data-driven change in mid-course. In essence, I get that the weight to the first half should be just under--let's see, I will go back and be really specific--the weight to the first half should be just about 45 percent. In fact, the

weight to the first half should be about 54 percent rather than 46. So, it turns out that you should give about a 1 to 1 rather than a 1 to 1.5.

The bottom line is it doesn't have a huge effect but it does have some effect on your estimate of survival. With a relative risk of 0.55 it becomes closer to 0.61. The alpha level as well has to be adjusted. By my count, the alpha level is probably closer toward 0.04 to 0.44 rather than the reported 0.012.

The other issue is that the information fraction is not as it was reported, 1050 over 1100 because the last 150 people largely weren't followed. You see that on page 16. You only see 3 of their events. So, the information fraction is closer to about 88 percent by my crude calculation, which means the O'Brien-Fleming with Lan-DeMets isn't as high as was reported, about 0.44; it is somewhat lower.

Consequences to all of this are not huge but they are not irrelevant. Instead of saying you are clearly across the boundary, I would say you

are right around the boundary, in fact possibly but not quite at it.

The question is do you have to do this adjustment at all? The adjustment, in theory, might have been said to have been made on the composite endpoint. Survival is a secondary endpoint. Well, the reality is survival is part of the primary endpoint and treatment effect on the primary endpoint is not independent of treatment effect on mortality. In fact, a principal reason that the primary endpoint looked better in the second half than in the first half was because mortality was dragging down the primary endpoint in the first half. So, I would argue you definitely do have to adjust for the fact that, even though the sample size adjustment was made on the primary endpoint, mortality is correlated to that primary endpoint and adjustment, therefore, has to be made.

The bottom line is the alternative to this is you design the trial initially. For example, if you were targeting a mortality size of a 40 percent reduction, you would have designed it to 110

events. If you design it to a 50 percent reduction, you design it to 55 events. That is roughly the contrast of how this study changed from a 65-event trial to 110. A much cleaner approach would have been to design the trial somewhat larger early and if initial results were large you would have early termination. Then the implementation of Lan and DeMets, O'Brien-Fleming monitoring guideline would have fully addressed these issues and there wouldn't have been any of these complexities.

DR. HIATT: Tom, can I follow up on that? That was what I was concerned about, the behavior of the DSMB. What struck me, from a more simplistic point of view, is that the first half of the study seemed to behave differently than the second half. I can understand biologically why that would occur. Now, I realize that the data went in the same direction so I guess if you did a test for heterogeneity it would still be a non-significant kind of thing.

But it doesn't make sense, and that is why

I was concerned that there could be bias from the DSMB at this one look to see that they are under-powered and then at the next look to stop prematurely. That just doesn't sound right. What you just explained makes perfect sense of why that is true. How we are supposed to interpret the data?

DR. FLEMING: When you monitor trials frequently one of the things that becomes apparent is that the early results often do reflect long-term later results, but they often don't. It is one of the major reasons why we shouldn't release early data. We constantly get re-taught the lesson that when early results become known to caregivers, or patients, or sponsors, or the investment community, prejudgment of those results occurs and impacts the integrity of the trial, and can also impact the efficiency and the reliability. The adaptive methods sound attractive--let's get in and then let's change our mind. They are not as efficient as thinking through the process at the beginning, setting things up correctly and then

not, in fact, putting yourself at risk to re-sizing things in the middle if they don't look the way you would hope that they should look because what you arthralgias seeing in the middle may not look the same. So, there are issues of reliability; there are issues of efficiency; and there are issues of integrity when you do that because you are, in fact, compromising the integrity if you are give indirect, if not direct, insights to people about what is happening early on.

DR. HIATT: Well, that is exactly right. And the concern I think is that in retrospect, hearing the DSMB story, it makes sense from a sort of narrowly defined patient safety issue--we have a signal here, why don't we stop?

But if you are now arguing the p value is 0.04 and the weight of evidence issue I think is on the table, then this is not overwhelming. Had the study continued I don't know if the groups would have separated further or not. But I think that question is begged by what happened. So, the behavior, in retrospect and my sense of it, was not

a good thing.

DR. NISSEN: You know, it is interesting. I am going to weigh in here a little bit and, from a non-statistician's viewpoint, there are several issues here in this trial that are a little bit different. First of all, this issue of having things bounce around a lot early in the trial. As you pointed out, Tom, there are innumerable examples where an endpoint has gone in the wrong direction early in the trial and then things turn right around, and by the end of the trial you have a highly significant result. If you do enough trials you are going to see that happen.

There are some special considerations here and we have to understand what they are. Very few trials have been done in a solely African American population. I am sure that the sponsor and everybody was uncomfortable whether they could find these patients and get them into the trial. I am sure this was a challenging trial to do. There is also the fact that this is a very small company and so there was conservation of resources, and

sometimes you make these compromises which, unfortunately, do impact on the strength of evidence. But I don't think, you know, with this agent that they were going to target for an African American population they could invest--and you would like to do a very powerful trial. In fact, you would like to do a trial that would be able to detect as little as a 15 percent outcome 20 percent reduction in mortality.

Let me tell you, as a clinician, you give me drug with statistically robust evidence that it reduces mortality by 15 percent, that is a blockbuster. This was under-powered from the very beginning and I think I understand some of the reasons why it was under-powered. The fortunate thing for the sponsor is the final effect size, this 43 percent reduction in mortality and pretty big effects in other areas that allowed the small sample size to work out for them. But, in fact, the fact is they were working on the margins of what an adequately powered study for this indication should look like, and those are

exigencies of trying to do this in a special population with a small company that has some limited resources, and I am sure that is exactly what happened here. Bob, you wanted to say something?

DR. TEMPLE: This relates to the last point Tom made and went over fairly quickly. The recalculation was not based on survival data. Actually, I can't tell--was it based on the original primary endpoint? Yes? Okay. That means that survival events were a tiny fraction of the total. I can see from page 15 that that they are only a third--that is only deaths and hospitalizations; that doesn't count the other ones. So, it is a tiny fraction.

You very quickly said that doesn't matter, you have to adjust the death anyway by the same amount. I am interested in that. It may be beyond the scope today, but that comes up a lot. We have told people that if you didn't do an interim analysis for survival you don't have to do O'Brien-Fleming on it because you didn't take a

look at it. In fact, they might not even have taken a look at it. That is another possibility. I guess here they probably did. But that seems relatively important and the adjustment you suggest for mortality is quite large. You are more than doubling the p value. In other words, even if you treated the data sets as completely independent you would only double them I think.

DR. FLEMING: It is part of the complexity--

DR. TEMPLE: That seems quite extreme for this particular thing.

DR. FLEMING: It is part of the complexity. But just to come back, Bob, you raise a really good point that I also have encountered frequently and think a lot about, and that is if a trial is looking at death, MI and stroke and it is monitored on that endpoint, when you go back do you have to adjust for an analysis that looks at death alone, or do you have to adjust for an analysis that looks at death/MI? Well, the death component is probably a fairly small part of the

death/MI/stroke. The death/MI component is a pretty substantial part so I think you can see right away that what matters is the correlation. What is the correlation bet the test statistic for the primary endpoint and the test statistic for the secondary endpoint? If, in fact, there is non-trivial correlation, then you are not stopping at random.

Furthermore, in this process when there is an adjustment to the sample size, can you assure me that the only thing people looked at was the primary endpoint and not at any of the components? But, furthermore, I would argue what led to the increase is a relatively unimpressive result that was particularly apparent with survival. It was the survival part that was particularly apparently different. So, the endpoint that gained the most, the endpoint in this analysis that gained the most from the increase in sample size was survival. If you look at the data here, Bob, when you look at the breakdown on page 15--

DR. TEMPLE: Hospitalization was not so

unimpressive.

DR. FLEMING: First heart failure hospitalization was 0.66, 0.58. The composite endpoint was 24 versus 31. Those types of changes themselves would coincide with 1.6 inflation of the sample size, which happened. But mortality was a 15-fold difference. So the very endpoint that actually was weighing in here to have this influence was particularly survival. It is the measure that benefited the most by the increase.

To say that I am not going to adjust for it--we can have long philosophical discussions but it is not controversial that you have to adjust. The question is do you have to take the full adjustment because it is correlated?

[Multi-member discussion]

DR. NISSEN: Implicitly you looked at everything.

DR. FLEMING: Implicitly, but even if you told me I only looked at the composite, the test statistic for the composite is correlated with the test statistic for survival. So, even there, there

is a feature of adjustment.

DR. TEMPLE: Even though it wasn't at the time.

DR. FLEMING: It is always correlated, Bob.

DR. TEMPLE: It wasn't in the first look.

DR. FLEMING: It is always correlated. Those two test statistics are always correlated. A test statistic on death/MI/stroke is always correlated with a test statistic on the components.

DR. NISSEN: Let me bring us back down to earth--

[Laughter]

DR. FLEMING: Maybe there is a bottom line here--

DR. NISSEN: Well, let me propose a bottom line and see how this sits with people, namely, we can view this almost as a sensitivity analysis and we can say that we have a best-case and a worst-case scenario. We could have the most liberal adjustment method, which is the sponsor's adjustment to 0.16 which, Tom, I understand you

reject as being not sufficient and that is perfectly proper. We have the most conservative adjustment--and I expect you to do this because you always have done this for us on the committee--which takes you up to as much as 0.44, let's say. So, the real answer probably is somewhere in that range and the committee may want to factor that thinking into their conclusions. Would you accept that judgment?

DR. FLEMING: It is not quite what you said because we have to distinguish the adjustments on the primary endpoint and what I was talking about was mortality. So, the mortality p value was 0.12 and I am putting forward what my best attempt was to recreate what the proper adjustment would be, after three hours of attempting to do this because of the complexity of what happens when you are doing these adaptive methods, and I come up with something in the range of 0.04.

The p values you are referring to are for the primary where the unadjusted was 0.01; the sponsor's adjustment was 0.16; the FDA's adjustment

was 0.021, just reflecting the fact that when you get into these adaptive methods there is a lot of confusion. You now have three different levels there.

What is the impact of this? It is not huge; it is not huge in this case but it is not irrelevant. So, if somebody says we have clearly established a survival effect crossing the boundary, I would step back and say we have clearly got evidence of a survival effect that is the vicinity of the boundary. Now, that is not a hugely different statement but it is actually somewhat different and it will have some implications later on when one looks at the overall strength of evidence.

DR. NISSEN: I think what you have heard is some statistical arguments about what is a more conservative approach to the data. I think it is entirely appropriate and the committee should, of course, factor in everybody's thoughts about that. John, do you want to say something?

DR. TEERLINK: I was just interested in

what is the boundary then. We all think anything below 0.05 is significant but actually when you do these kind of adjustments for boundaries 0.05 isn't the target anymore. There is a new target when you do these. Because of what we have seen in trials where it bounces around, we have seen trials that were wildly positive and turned negative later, and the reverse also. So, to protect against that random chance in the swing, especially when you are dealing with such small numbers of events, you have to readjust what that p is that is the target. So, what was that target?

DR. FLEMING: Well, in my view, to come back to what I was saying before, the p values that I truly best understand are the prespecified primary analysis and the prespecified primary endpoint. I am extending that to saying the sequential monitoring of those p values, using a proper adjustment, and I believe this trial hit that endpoint. It is statistically significant on that particular measure.

DR. NISSEN: But barely so.

DR. FLEMING: Well, let's go with the FDA adjustment, 0.021. That is, in fact, below the traditional 0.05. Now, a big issue that should come up in our minds is if this is a single trial registrational package, what is the target? Ray Lipicky used to remind us if we think of one-sided p values of a positive trial as 0.025, he used to say what is 0.025 squared, which is triple 0.625. And, there have been many discussions that if you have a measure like mortality--let's just go back to January 6th and 7th of 2003 rather than recreate wheels--what if mortality is, say, a secondary measure, is different from other measures, shouldn't there in fact be some particular attention given to mortality? I would argue yes, but it is not free. It is the secondary endpoint. I would still want to see some greater strength of evidence.

Back of the envelope two years ago, I said put zero in front of it; 0.0025 one-sided would be persuasive. But all of that still is strength of evidence of a single trial. If a trial should

stand alone, both FDA and European regulatory authorities have said, yes, single trials can be an adequate basis for approval. Results have to be robust and compelling, pristine trials, internal consistency, high quality conduct and, vaguely stated, stronger statistical evidence.

Frequently across many disease areas, many divisions in FDA have heard that stated as if it is an irreversible morbidity or mortality endpoint, like death, like stroke, like loss of vision, like HIV infection, something less than strength of evidence of two trials would be a basis. Well, I am a statistician--what is a number between 1 and 2? It is 1.5. So, 0.025 is 0.04; two-sided p value, 0.01.

So, back of the envelope that I have always used on a primary mortality endpoint two-sided 0.05 is SOE-1. Two-sided 0.01 is approvable; 0.01 is approvable for a non-mortality measure. Purely a guide, factoring in other things like secondary endpoints--safety, etc.

So, just to get to the essence here, for

mortality I think it is close to what I would call SOE-1, close to 0.025 one-sided. It is in that vicinity. The question though is what is the strength of evidence that is going to be required for that, given that mortality wasn't the primary endpoint?

DR. NISSEN: I had a very similar discussion in a discussion of pulmonary hypertension where we were dealing with something that was kind of an orphan disease, and I remember, you know, we sort of said, well, you make some adjustments sometimes because you want to encourage trials in special populations and diseases which are of public health importance which we have few therapies for. I just want the committee to think about this.

I am not offering an opinion, although I have one, that with this problem of not having trials in minority populations, of the disproportionate burden, the public health consequences we sometimes, as a committee, have made adjustments based upon factors that are not

statistical. We have said many times on this committee that we are never going to be slave to a p value. We are here to do the most good for the most people. So, we will have to factor in--you will all have to make your choices, as will I, about that. It might as well be explicit that, you know, we are thinking about those things. Norman?

DR. FLEMING: Let me just add to that though. As a statistician on the committee I strongly endorse that. My interest is to do the best we can to try to put into proper context what is the statistical strength of evidence, but then clearly encourage that a lot of clinical judgment needs to be brought in, in terms of the totality of the picture, what is known from external sources, how extraordinary is this instance, etc. All that definitely needs to be factored in.

DR. NISSEN: Norman, did you want to say something?

DR. STOCKBRIDGE: It was my hope to get you to be explicit about your thinking about the earlier studies in this. To what extent does that

reduce the burden on the third trial? This isn't, you know, just dreaming outside of anything, sitting down to do a single trial. Do you give any credit at all for the hypothesis that was generated out of the first two?

DR. NISSEN: You will hear from us about that, I have no doubt, as we move forward.

DR. OTA WANG: I actually have an additional comment.

DR. NISSEN: Please.

DR. OTA WANG: There has been talk that because there are certain populations or communities that are difficult to recruit from or there are limited resources, I am uncomfortable with the thought that there is a notion that for some types of research, for some types of communities or populations we can actually lower the bar in terms of scientific integrity that we are using to evaluate the research. So, I guess I would also like to bring up to the committee to put that into weighing the evidence. Should we actually have different standards for different

types of research for different communities? And is it justifiable to say that we can actually do that because there are limited resources or that because some people find it difficult to actually enter those communities to recruit them for subjects?

DR. NISSEN: I am going to answer that for myself and tell you that, you know, if you develop a drug and the people you can enroll in a clinical trial is the entirety of the population in the United States for the disease, let's say heart failure, it is a lot easier to study that population than it is a population that represents a relatively small fraction of the population. This is just a practical matter. We love to have trials that have more than enough power to answer the questions very, very well. That is hard to achieve when the population that you are trying to study--the FDA has recognized this in some of the policies related to orphan drugs where you have a small number of people that have a disease. So, if you are developing a drug for a disease and there

are not many people that have it, you get some points for doing that.

I am arguing that it is not unreasonable public policy to make some adjustment for that. Now, you may not all agree with me, and Bob Temple may not agree with me, but that is one guy's opinion.

DR. TEMPLE: I might agree under some circumstances. I don't think it comes up here. Despite what some people think, the enrollment of the black population into cardiovascular trials is not so miserable as people think. It is not 5 percent. I will circulate a document we did some years ago. I don't think the data has changed too much. It is more like 15-30 percent, perhaps because there is a high rate of cardiovascular disease.

So, I don't think there is any difficulty in finding a suitable number of blacks to enter a study like this. I don't think we should make any particular allowance for that. It ought to be good data and it ought to meet the test.

There are rare circumstances where, you know, there are only ten people with the disease. That is another kind of question. I don't think that is the issue here. I really don't. There is a lot of black participation. Actually, for what it is worth, we are having trouble now because trials are being done abroad; it is getting worse.

DR. NISSEN: Let me just say I am not saying how large the adjustment ought to be. You know a p value of 0.15 is not what we are talking about here. What we are talking about here is where something is close to what we would consider to be compelling. I think we all agree that we are in that sort of ballpark. Do you get points for doing a study in people for which it is more difficult to do the study and where the information is very valuable from a societal point of view and from a medical care point of view?

You know, I live in Cleveland, Ohio, Bob. We have a very large African American population. We see a lot of heart failure. As we all know and we are to talk about a little bit later, ACE

inhibitors don't work so well in that population. So, when you get information that is potentially very valuable and informative about a group that can be very difficult to treat, you have to give a sponsor some points for going after that.

DR. TEMPLE: I agree. You just said a different thing though. This is new information in a group that is under-served by available therapy. That is a different question from a bunch of people hard to get into trials.

DR. NISSEN: Well, I am factoring all of those factors in. You know, Tom is very good at adjusting p values for statistical considerations. My job as a clinician on the panel is to adjust p values for clinical and I think societal considerations, and I am going to tell you what I think a little bit later.

DR. TEMPLE: Well, it even goes to the importance of the finding, and it is no secret that we treat death differently from the way we treat other things in terms of adjustment and that makes a lot of sense. So, I don't object to any of those

things. I just want to tell you it is not that hard historically to get the black population into cardiovascular trials.

DR. NISSEN: You know, it is hard for some of the reasons you heard from the microphone today. That is, there is distrust of the health provider community by African Americans, some of it justified. So, we have to overcompensate in order to make people comfortable in minority groups with participating in clinical trials. Now, we are working at that and we are doing a lot of work to try to do that. These folks were able to pull it off and I am going to give them some points for that. Let's move on.

MR. SAMUELS: Mr. Chairman, if I may? You know, I have been sitting here and listening. I am neither a clinician nor am I a physician. I am here as a patient representative. It has been fascinating for me because for the last ten years my experience has been primarily in the world of cancer. I am an 11-year prostate cancer survivor and a 6-year throat cancer survivor. I also happen

to have high blood pressure and diabetes. So, I think I am the audience that this drug is really meant for. And it has been fascinating to hear the statistical perspective, the clinical perspective. But let me tell you from a patient perspective.

I hear the statistics of people who have been impacted by this disease and, indeed, I probably will be one of those who will suffer from it. I would hope that we would understand the fact that we may have to adjust slightly to accommodate the need of this very hard-pressed segment of our population that suffers disproportionately from cancer, from heart disease, from diabetes. I mean, there are just so many health issues that affect this community. So, I have come with my mind not made up but I will tell you that after listening, the people who mean the most to me are the patients because, indeed, that is truly who this is going to affect.

It is kind of hard for me to justify from a statistical/clinician point of view the validity of all that we have heard today. But from a very

practical point of view, as one who, indeed, will probably wind up having to take this drug or one like it, I can only urge you to think like a patient. Please think like a patient because, indeed, as we walk away today this is a historical decision because, as I understand, it is one of the first drugs that will be focused at a specific segment of our population. So, I can understand the discussion and debate and the hardness of this decision, but I just ask if you could think in your heart as a patient when you ultimately make your decision about this.

DR. NISSEN: thank you very much, sir. I am going to move back to the questions. Obviously, this is a very challenging and precedent setting meeting. I believe we are on 1.2.2. The question is how comforted are you by analyses of CHF hospitalization among early and late enrollees? Have we addressed that adequately? I think we have.

1.3. The difference in time to first hospitalization for heart failure was large and

statistically significant, while the difference in total days in hospital for heart failure or for other cardiovascular causes was small and statistically insignificant. For patients with heart failure, is time--

DR. STOCKBRIDGE: I think we can skip that one. I think the company adequately addressed that during the presentation.

DR. NISSEN: Okay. Do you want to skip all of 1.3?

DR. STOCKBRIDGE: I think we can do that, unless somebody thinks they didn't address that.

DR. NISSEN: I thought it was nicely addressed also. 1.4. Interpretation of the quality of life data is rendered difficult because of the early termination of the study. How persuasive is the retrospective analysis with last observation carried forward? Anybody want to jump in on that?

I am going to say a few things here. First of all, it is the most difficult but it is also very important data. The first thing I want

to say is that I am glad that it was included. I am not sure if including it as a part of the composite primary endpoint was the wisest approach as opposed to a well-defined secondary endpoint, but what was done was done.

The question really I think speaks to the robustness of the data. It is clearly harmed significantly by early termination and that is why Tom Fleming's concerns about the exact circumstances under which termination was decided are very important. Because whenever you terminate a trial early you take away information and that information can sometimes be extraordinarily valuable.

Having said that, in interpreting the data there are several things that help me here. I actually really liked the slide where we got to see the point estimates for each of the time points, 3 months, 6 months, 9 months, 12 months, etc. It is very helpful that you see that at virtually every assessment, sometimes significant, sometimes not, things are going in the right direction and by

approximately the same amount. I would consider the last observation carried forward versus the initial analysis to be a sensitivity analysis. So, no matter how you look at it, you end up with a p value, no matter how you slice it and dice it.

I also think that for future trials, having seen what happened here, I am not sure I would pick a single point in time to do this because I think it could be very distorting in other trials where things can bounce around a certain amount, and maybe there is a more robust way to do that. Having said that, the consistency of the fact, no matter how analyzed, I felt convincing. Tom?

DR. FLEMING: I largely agree. When I think in terms of endpoints in trials, particularly primary endpoints, I think of being guided by the principle that I want that endpoint to be measurable and interpretable. I want it to be sensitive to the effects of the intervention. And I want it to be clinically relevant. To the sponsor's credit, I think they have chosen

endpoints that are clinically relevant. They are putting together mortality; they are putting together heart failure, hospitalization; they are putting together quality of life.

I think in the interest of maximizing sensitivity in a single measure they have given up some interpretability, and it is exactly the point you are making. I would have preferred to have had heart failure and hospitalization-free survival, i.e., those two components together, heart failure hospitalization-free survival and a quality of life assessment either as co-primary or as a secondary because it greatly enhances the interpretability. In fact, that is the way I am interpreting the data. I get lost in this primary endpoint. I know that we like a single measure and I like a single measure too but sometimes we give up too much interpretability in that goal.

So, as we look at the component of quality of life, my sense is similar to yours. I find this reassuring. Do I call it statistically significant or not? I am a little uneasy with this LOCF, as

you can tell. But the main drift of this that I am coming away from is, yes, this is very reinforcing to the other components, to the favorable pattern seen on heart failure hospitalization and on overall mortality. There are certainly indications of overall benefit in quality of life, although I wouldn't want to be pinned down on exactly what strength of evidence that would be because of these uncertainties of the missingness, even if a fair amount of that missingness was because of early termination.

DR. HIATT: From a clinical perspective on that, I agree with your interpretation. I think that the sequential nature of the group differences strengthens any of the sort of how to carry forward data or impute the worst score. But the scale around that instrument versus a hard endpoint is so hard to interpret in the context of a bundled endpoint.

I know for the SF-36 physical function components there is a certain expected change in the healthy population over time. They lose a

certain amount of function per year and an intervention that changes that measurement by a certain amount gives you back a year of functional life, if you will. It is not really quality life but it is certainly function. I don't know as much about this instrument. Obviously, a lot of people around the table do.

I guess what I am asking here is what those group differences mean in terms of the patient? Is going from Class III to Class II? I don't know if we need to debate that today but I think it is harder to interpret in the context of a bundled endpoint. I would have rather seen it as a secondary.

DR. NISSEN: My colleagues in the heart failure world whom I do consult from time to time tell me that these differences have been shown and been validated to be important differences in terms of what they really mean to patients.

I am going to take a little bit of a contrary view on something. That is, some people would argue that the purpose of modern medicine is

not to extend life; it is to improve the quality of life. You could even argue that if you had a drug--suppose you had a drug--people with heart failure are miserable so if you had a drug that had no effect on survival, no effect on hospitalization and a robust effect on quality of life it might be a pretty good drug. You know, if you talk to anybody with heart failure they will tell you not being able to breathe is one of the scariest things that human beings go through. This is important to me, Tom, as a clinician, as somebody who has to look across the exam table at a patient and say can I make you better? Can I made you feel better? So, "feel good" endpoints are important endpoints and this is a "feel good" endpoint and I value it.

DR. FLEMING: I remember sitting on the oncology drugs advisory committee in 1985, sitting across from Bob Temple--

DR. TEMPLE: What month?

DR. FLEMING: I don't remember, Bob, I am sorry.

[Laughter]

In essence, the critical discussion at that meeting was exactly the point that you were making, Steve, and that was, we were sitting around saying we may not in some cases be able to prolong survival of cancer patients but if we could meaningfully improve their quality of life that is a great advance. For 20 years we have been struggling with coming up with sensitive, validated measures of quality of life, knowing what meaningful impact is on those measures, and we have made some progress. But, in essence, the challenge has frequently been not to recognize that it is important to improve quality of life as well as survival, but to be able to do so in a conclusive, persuasive way.

DR. TEMPLE: I take it no one remembers that the cardiorenal advisory committee has opined on this matter when it was considering flosequinan. For all I know, either Jay or Milton was the chair at the time, I don't know. There was a point estimate for survival that was adverse. This was before a profile was done. The committee

explicitly said if you can show improved exercise tolerance we don't care. It is worth it anyway. When the profile came out, which is in abstract form and has never been published--

[Laughter]

--but we all know the results, it showed not only that survival was impaired but it showed that exercise improvement was gone by three months or so. So, that was a convincing case for getting rid of the drug. But the committee really did think about this and agonized about it and offered that opinion.

DR. NISSEN: I want to put that on the table and obviously it is going to come up again, and we might as well sort of give our thoughts about it. I am going to move us forward unless there are further comments about this. Yes, Jonathan?

DR. SACKNER-BERNSTEIN: For the most part I agree with what has been said about the quality of life as an independent versus as part of the composite. But the one concern that I do have is

that we are looking at a therapy that is in theory based on some biologic difference between blacks and non-blacks. In fact, there are probably more physiologic differences between men and women than there is between blacks and non-blacks. So, I am a little bit concerned that we really haven't gotten into the issue of how little data there is--even though the point estimates are favorable, how little data there is for women to be treated with this. There is no external data to corroborate it. So, even though the point estimate is small and the confidence intervals don't have a lot of unity certainly with mortality--you are talking probably only about 300 women that saw this drug for six months.

DR. NISSEN: Stay with us because we have an upcoming question where we are going to ask about the population--

DR. SACKNER-BERNSTEIN: The question I was going to ask is should we find out about quality of life in women specifically? That is why I was bringing it up now. Do you want to wait till we

get to the population?

DR. NISSEN: Let's do it when we get down to that topic. I am going to try to move us right on through the questions. We made some progress there.

Number 2. Policy issues. Ordinarily, one expects to understand the role of each component in a combination product, as noted in 21 CFR 300.50. How important would that be if you believed there was an effect on mortality? On hospitalization? On symptoms? If there had ben more than two active ingredients and if you suspected one component is subject to tolerance effects?

I will take all those together, and what we are really talking about is this issue of do we understand the components and this is not the usual approach. So, let's hear what the committee has to say. Anybody? Go ahead, John.

DR. TEERLINK: I think it is fairly clear we don't understand what all the components are doing. There is no information on dose. There is no information on how to adjust either component

for any effect or for any positive outcome or for any adverse effect. This is the only trial that has been done with this dose. This is the only information we have with this pill.

DR. NISSEN: I am going to make some comments that I think might be relevant to what you are asking here. First of all, you said it earlier, Bob Temple, and I agree with that, if you believe that there is an effect on mortality there is just nothing you can do here about this because, as you pointed out rightly earlier, you can't do a factorial study when you expect that the mortality advantage is coming from one component. You can't deny people that component.

Now the question is, is the evidence on mortality persuasive enough? And, you know, apart from the p values--you know, a 43 percent reduction in mortality is a pretty big effect. If you think that is coming from one of the components, in order to find it out you are going to have to expose a group of people to a pretty substantial risk.

The question would be more interesting, of

course, if it were on softer endpoints, such as hospitalization and relief of symptoms, etc. Again, this is I think potentially agonizingly difficult because our job is to avoid suffering, but when you do clinical trials you sometimes--well, you are always going to have a placebo. Right? So, you do a placebo-controlled trial. If the drug actually works and if you really think it is going to work, you are denying half the participants the benefit. Why is there equipoise? There is equipoise because we don't know. If we don't know for certain, then the trial is ethical.

So, the question is where does that boundary lie? That is a very, very difficult question. I am not necessarily prepared to answer it sort of in the abstract. I think the answer to that question is in the specific case. So, let's take this specific case. I guess I think that given the fact that we are talking about legacy products that have always been given together, going back a long way, there are priors here that

have some influence on our thinking about this. I guess I think that the fact that you have a positive trial now really does pretty much preclude taking it apart and figuring out which component is leading to which aspect of protection. But somebody else may not agree with me.

DR. HIATT: Steve, I think I support that. I mean, I raised this issue earlier today and I was quite concerned about the lack of information on the components and the doses. I still think that is a limitation. But I think the argument you just made trumps that and it is difficult to argue against it.

So, I would then like to suggest--and I can't predict how the vote will go right now--but if this were to go forward, I really think it is imperative that the potentially rare toxicities be formally and rigorously evaluated and not just left to some open surveillance kind of mechanism. I think the issue now is not does this drug maybe lower mortality risk because we think we know that in the current formulation, the current dose. The

issue is, is there potential toxicity to a component? If there is, how are we going to detect that? So, I really would like the FDA and the committee to consider that.

DR. TEMPLE: You are particularly worried, I take it, about lupus developing over time?

DR. HIATT: What I am worried about is that hydralazine was given a long time ago. It is really not given a lot very much now. It will potentially be given to a lot of patients that haven't been exposed to it before. Jonathan mentioned the fact that a lot of women might be taking the hydralazine who might be at higher risk. I think lupus is only one example of drug toxicity that has a small signal and that wouldn't be detected in a clinical trial of this size. So, we may never know until something bad happens and I don't think today we can afford to be in that position.

DR. TEMPLE: But there are various ways of trying to get postmarketing information. It is easier sometimes if you are focusing on a

particular thing but you are not. You are worried about a more general issue, and largely because the use of this drug was a long time ago. I mean, there was very widespread use of hydralazine but it wasn't in the modern era with modern warning systems or anything like that.

DR. HIATT: I think it is a more generic issue around postmarketing surveillance that has been well reviewed in the medical literature in the last couple of months, and I also think we should be mindful of that today.

DR. NISSEN: I am going to take a contrary viewpoint here. First of all, I always think it is good to do postmarketing surveillance but it is important to put this in a certain clinical context. We know that there is some very low background rate of SLE-like phenomena with the use of hydralazine. We are treating a disease where the 2-year mortality is pretty high. You know, we have all been sensitized by the Cox-2 debate and some of these other things where we have a drug that was used for a non-life-threatening disease

and where a life-threatening toxicity occurred. Here is a situation where we have a life-threatening disease and a non-life-threatening toxicity. So, my threshold, my worry level is much lower when you are using a drug to treat--I don't know, what do you estimate, John, the 5-year mortality now is in Class III, Call IV heart failure? Fifty percent probably? Higher?

DR. SACKNER-BERNSTEIN: Even if it were less--

DR. NISSEN: Yes, it is still enormous. So, we are talking about a disease which is not inconsequential but, you know, lupus is not as serious. So, my concern is not as great but I understand where you are coming from and I do think drug safety is very, very important but is less important when you have a life-saving indication.

DR. TEMPLE: I guess the thing I would add as part of the general discussion is that the spontaneous reporting system or looking in certain healthcare systems is pretty good at picking up bolts from the blue type events--lupus, things like

that but, of course, it is not very good at detecting a 30 percent increase in heart attack rate. The only thing that gets that is controlled trials. So, you have to figure out what it is you would want to look for and whether you need more than the spontaneous reporting system in this particular case. I guess Steven is saying he thinks that will pick up what you need to know.

DR. NISSEN: That is sort of what I am saying.

DR. TEMPLE: It is also true there are increasingly ways of looking within captive data systems that allow us to look for those things that we are actively working on.

DR. NISSEN: Norman?

DR. STOCKBRIDGE: Before giving up totally on the thrust of 2.1., let me go around a different direction on it. I have over the last 12 months had several development programs presented, prospective development programs presented to the division that involve two, three, four, five different chemical entities. All right?

Prospective-- don't know whether or not it is going to affect mortality; don't know whether it is going to affect hospitalization, something you think is important. What am I supposed to tell the next one that comes in? Do they have an obligation to work up the contributions of the individual components? Or, do we wait and see whether or not it affects something you care about and then make that decision?

DR. FLEMING: Norman, I have been sitting here thinking about exactly that as this discussion has been ongoing. I am comfortable with all that has been said, and I think, for reasons that make sense, this question is being answered in the context of what we should do based on where we are today and what we should require for BiDil based on where we are today.

What concerns me is that there might be an interpretation of this discussion that, while it isn't so critical to understand what the components contribute and I would strongly disagree with that, I think if we believed in this setting that it is

very plausible that the essence of benefit to risk could have been achieved with any single component, then I think this points out the fact that this is a discussion that should have been held before the trial was done so that we would, in fact, design the right trial to be able to get us to where we want to be before we had data that would or would not indicate that there was a substantial survival effect.

So, to my way of thinking that doesn't mean there is a single right way to proceed. How strong is the evidence that the benefit you are going to get is truly through a synergy of these two components that really is inherently going to require both, well, then prospectively if that is persuasive I can accept going forward with the trial design that does control against both. But if there is far less insight than that and you have two classes of agents or two components that readily individually could carry a substantial essence of benefit to risk, then some type of a factorial design or other design that would

prospectively allow you to determine that so that when you had the data in hand you wouldn't be in the position of saying, well, we really don't know but now can we step back and readdress the question.

DR. NISSEN: Norman, my answer is similar to Tom's, with a wrinkle here. I think you have two sides to the equation. What is the evidence that combining the components is a rational approach to the disease state you are treating, that you have some priors; you know some things that make you think that that is a smart thing to do? Then, what are the toxicities that we know, or what do we know about toxicities? Because we all know--certainly I know that the more drugs you give a patient the better the probability that you are going to produce an adverse event. Polypharmacy leads to adverse events. So, if you can get something done with a single drug, that often will be the preferred approach.

So, I think what I would suggest that you want to do is look at what you know about whether

there is a reasonable basis for combining the agents; look at what you know that relates to the potential additive toxicities involved in doing so; and then on a case-by-case basis try to do what makes reasonable regulatory sense.

In other words, if you have weak evidence that combining the drugs makes a rational sense and one or more of those components has a high risk of toxicity, it is mandatory to test this in a factorial sort of design.

DR. HIATT: Well, it may be that the indication matters too.

DR. NISSEN: It does.

DR. HIATT: So, I think what we are talking about here is that if mortality is driving this it is unethical to try to figure that out. If the indication is a symptomatic endpoint it may not be unethical to figure that out.

DR. TEMPLE: I don't think we have backed off or are prepared to back off from the general requirement that you have evidence that both contribute. Now, what that evidence is could be

debated in any given case. My guess that a pure arterial dilator is probably not going to do so great in heart failure. Prazosin doesn't. Calcium channel blockers don't. They are all good arterial dilators. So, there is a lot of reason to think this makes sense. We may be able to get some more hemodynamic data together even after approval if you recommend it and we approve it. So, we can explore that. But I don't believe we think the combination policy or rule is in doubt. We expect that kind of information. That is what effectiveness means for a combination. But there are those special cases where you have a survival effect where it gets very dicey.

DR. NISSEN: It gets particularly dicey now if you have three or four or five components.

DR. TEMPLE: We are actually rewriting our combination policy and it will address things like that. We have had people come to us with 20 components, saying, okay, what am I supposed to do? And we have said show that improves survival and we will talk.

DR. MCCLESKEY: Bob, may I say something? From an industry perspective, what you just said is refreshing and in response to the question Norman raised, and that is that I think industry is pleased to hear that factorial designs that we are talking about here are not absolutely mandated but will be judged to their need, and so forth, on a case-by-case basis. That is refreshing to us.

Secondly, regarding the postmarketing safety issue that you discussed before, in the specific case I wonder if it is worth considering what we have heard from testimony in front of this committee that this particular patient population is sometimes difficult to have access to; difficult to follow-up; and some kind of postmarketing safety evaluation of that group might be more difficult than of another patient population.

DR. TEMPLE: You know, I don't know that. I don't know if there are fewer adverse reaction reports or not. If you wanted to do something special, you might consider a registry. Registries are voluntary. If you don't want to do it, don't

sign up. And you might be able to do that. I think Steven addressed the question of whether that is necessary.

Just on the first part, don't be over-refreshed--

[Laughter]

--we really would ordinarily expect factorial designs unless someone can make a convincing case that it is obvious from other source of data. That is allowed, always has been.

DR. NISSEN: You know, this is not irrelevant. There are some strange folks running around in the U.K. that want to do something called a poly-pill and they are going to throw about seven things in it. Who knows, it may come before you some day, or before us.

We are 2.1.2. What is the evidence that both components heart failure hemodynamic effects when used together short term and long-term? Do you really care about that? You know, the hemodynamic effects are not really the basis for approval, are they? Tell me what you are looking

for there.

DR. STOCKBRIDGE: I was just looking for a hint somewhere that you really did need both components here. Do you think you have a hint that you need both components?

DR. NISSEN: Anybody want to comment? I have my own thoughts. You know, there is a reason why Dr. Cohn and his colleagues originally combined hydralazine and nitrates. From day one these components were used together and the reason was very simple. There was hemodynamic data. If you put a right heart catheter in patients and you gave hydralazine the cardiac output went up but pulmonary capillary wedge pressure did not fall. If you gave nitrates pulmonary capillary wedge pressure fell but you didn't see a lot of fall in arterial resistance. You had one drug working on the arterial side, one drug working on the venous side. It was a logical thing to do and pretty smart given when it happened, which is a long time ago. Somebody thought of the idea of putting the two of them together and giving it to heart failure

patients.

Now, we know hemodynamically that that, in fact, is the case. Whether that is the reason why the combination led to the results that we have here, I can't vouch for that but that is the rationale that existed. I was sort of around during that era.

What instructions do you give for patients who do not tolerate one component of BiDil?

DR. SACKNER-BERNSTEIN: You wouldn't be able to tell if they didn't tolerate one component of BiDil based on this data.

DR. NISSEN: Yes and no.

DR. SACKNER-BERNSTEIN: Headache can be produced by both; hypertension by both; dizziness by both.

DR. NISSEN: Yes.

DR. SACKNER-BERNSTEIN: So, lupus syndrome, a peradoxime associated neuropathy, peradoxime responsive neuropathy, those are associated with hydralazine. Angioedema, I guess I would be suspicious that it is the hydralazine that

is causing it and not the nitrates. But aside from that, I think it is very hard to tell.

DR. NISSEN: There is another question they are asking here, of course, which is that since the drugs are available independently--you know, did anybody who has ever had a patient with a nitrate headache get a really bad one? You know, it is pretty distinctive. It is a throbbing, nasty headache. Would you go ahead and separate the drug? I can tell you what I would probably do as a clinician. If I had seen a very good clinical response I would probably stop the BiDil, put them on hydralazine and nitrates and back-titrate the nitrate.

We didn't talk about this, but there is an issue with oral nitrates. We might as well have at least a little discussion of this. Oral nitrates are subject to first-pass hepatic metabolism. As a consequence, the range of doses at which you see toxicity from oral nitrates is very broad, and anybody who has ever used those drugs knows that. You can give people 5 mg of isosorbide dinitrate

and get the worst splitting headache you ever saw in your life. In other people you can throw nitrates at them until you are blue in the face and you don't seem to get the toxicities.

So, this issue of drugs with first-pass metabolism--these are drugs that we often like to give, watch the relationship between the dose we give and the effect that we see and do the titration. There may be clinicians that will choose not to use BiDil. They will choose to use hydralazine as a separate component and isosorbide dinitrate as a separate component and then individually titrate them. We are not precluding anybody, if we approve this, from doing that.

DR. HIATT: But, Steve, I am not sure it is logical because the outcome data are based on a fixed dose combination. We also know there is tolerance to nitrates as you push the dose. You may lose the benefit of the nitrate effect if you push the dose. We don't know that.

DR. NISSEN: Well, you are arguing against my strategy but I might argue with you that if I

have an African American patient that has max'd out on everything else, I might just make a clinical decision that giving him half as much nitrate with a full dose of hydralazine is better than giving him nothing at all, and that is okay. That is a clinical judgment that physicians are perfectly inclined to make. That is not a regulatory issue. Those are approved drugs. I can use them in any way I want.

What you are arguing is it might not meet the strength of regulatory evidence but it might meet the strength of evidence when dealing with an individual patient. We always can do that. That is one of the nice things about what we are looking at here. You can always go back to the old way of doing this if you choose to.

DR. TEMPLE: But what I heard is that nobody thought there was any advice we were capable of giving in labeling at this point.

DR. NISSEN: I don't think so.

DR. SACKNER-BERNSTEIN: As I understood the documents, the protocol seemed to give

investigators the idea that if there were side effects, presumably hypotensive type side effects, the preexisting therapies should be adjusted to maintain active therapy, if I interpreted the documents right.

DR. NISSEN: In clinical practice in this area, I don't think you can be very specific, and I think that the need to individualize therapy in heart failure is well recognized and the fact that a lot of the drugs we give can produce hypotension. Physicians will have to make their minds up about which of the components they will down-titrate if they get, for example, symptomatic hypotension. I would be very cautious about offering that sort of advice. It is just not probably what we ought to be doing from a regulatory point of view--one man's opinion.

Moving along, ordinarily, one expects to know something about the effect of dose, and one does not in this case, for either component. What does the importance of information on dose change? With the endpoint? With the number of active

ingredients? Anybody want to say anything about that? Norman, what are you looking for here?

DR. STOCKBRIDGE: Well, I am just trying to get some sense of how important knowing something about dose ever is.

DR. NISSEN: You know, it is wonderful when you have it. We don't. I think that it is like everything else in clinical medicine. You know, you have this wish-list of things you wish you knew about every drug. You should have absolutely elegant information on the dose-response curve for every drug we use and for some of them we do. We know what happens as you raise the dose of antihypertensive agents. You sometimes tell us that in labels. But often for drugs we don't really know that data. Look at the other heart failure components that we use, we don't really know a lot about that. We know what the average dose was of enalapril in the consensus trial. We know what the average doses are in trials. We often cannot figure out the information about whether that makes a difference or not, and it is a

confounder that physicians have to deal with every day.

DR. STOCKBRIDGE: Suppose you didn't think there was a mortality effect here, suppose it had come out pretty even with mortality and you thought this was a hospitalization benefit, the composite still won so you thought the thing was interpretable. Do you now begin to care about dose and want it worked up?

DR. HIATT: That is relevant to what we just said I think, isn't it? The argument that trumps not doing that is the one that Steve proposed. It is the mortality. If you are suggesting that it is more of a symptomatic benefit, i.e., hospitalization, then you probably have more license to explore does.

DR. STOCKBRIDGE: Well, it is not license, it is encouragement. Then, the other part of this question is suppose there are more than two components here. How does that influence your interest in knowing something about that?

DR. NISSEN: It makes it much more

difficult. I think now the number of combinations you are looking at here really goes up exponentially. Again, there are some circumstances where you have a treatment effect that is very easily observed and measured where such studies can be done. But it is very hard in longer-term trials of drugs in these sort of life-threatening conditions to get the most elegant information you would like. Ray Lipicky used to sit here and argue that he wants you to go beyond the best dose and find the bad dose, or whatever, the dose where you no longer have any additional effect. Certainly, from a scientific point of view that is often very useful information but it is very hard to achieve.

DR. TEMPLE: Where the number of people in trial group is manageable we, in fact, do that. Combination studies for antihypertensives have been carried out in which there were four doses in one drug and three doses in the other and you get this nice response surface, and you really do know what the dose of each is. But that is where you can get away with doing 30, 40 people per group. In

outcome studies it is very unusual to have more than one dose in an outcome study. That is a significant problem sometimes but it is not clear what to do about it. We sometimes actually have been encouraging people to start at low dose and then drop it if it doesn't have some advantage on something, and then go to a high dose. But you still don't really know that you wouldn't have done as well with the low dose. You just know that it wasn't toxic.

DR. NISSEN: There may be therapies out there we use every day where if we gave twice as much of the drug we would get twice a much effect, and we will never know because it is not likely to get tested. But, you know, that is a problem and that is a problem that is extremely difficult to solve. Unfortunately, when you are looking at morbidity and mortality trials where you have serious morbidity and mortality it is really tough to do it.

DR. TEMPLE: Not only that, it is hard enough to show a difference between the treatment

and placebo, now you are trying to show a difference between two treatments, both of which may be active but not quite as active and you run into sample sizes that are really quite daunting. It is a problem.

DR. NISSEN: Believe me, I know that because I do it for a living in the cholesterol world where we have done some active control trials and they are tough to do.

Subjects randomized to BiDil had lower blood pressure than those randomized to placebo. Is this a plausible explanation for the differences in outcome? What should labeling say about observed differences in blood pressure? Great question.

DR. HIATT: There is obviously dissociation there. I mean there are drugs that lower blood pressure--we have already discussed prazosin--that don't improve outcomes. Beta-blockers may not lower it as much. I am not sure the answer to 2.3.1. is that that explains the benefit. It may be part of the mechanism but we

can't say that definitively.

DR. SACKNER-BERNSTEIN: I think that if the BiDil patients that had significantly higher blood pressure at baseline, it would be a more important question to ask because then you would wonder whether this was almost an antihypertensive unloading effect just by selecting a different population. But it doesn't change how I view things.

DR. NISSEN: I am actually very interested in hemodynamics and we don't want to forget about the fact that blood pressure equals cardiac output times systemic vascular resistance. So, if you give a drug that lowers systemic vascular resistance, which we think hydralazine does, and if it is in a heart failure population, then cardiac output goes up and blood pressure may actually stay the same. So, there tends to be this problem that because blood pressure had two components, resistance and output, one of the therapies you are giving is altering output and, therefore, you don't really know. In fact, we occasionally see a

patient that we put on an infusion of sodium nitroprusside, one of the most potent vasodilators in the world, and the blood pressure goes up as we are up-titrating the nitroprusside. Our fellows are very surprised by that but I am never surprised by it because I understand the physiology. Yes, Tom?

DR. FLEMING: We ought to have a meeting to talk about whether blood pressure is important!

[Laughter]

If we go to page 28 in the FDA review, medical officer review, I thought the indication there was that the blood pressures at baseline was higher in the BiDil group, 128 systolic against 124 and diastolic 77 and 74. So, I am a little confused. The question says BiDil had lower blood pressure.

DR. STOCKBRIDGE: No, he was talking about during the trial.

DR. NISSEN: This is baseline. Tom, what he is trying to get at is would any antihypertensive drug, added to this regimen that

patients were on, produce the same effect. We can't answer that. I know you know we can't answer that.

DR. FLEMING: I am with you completely and the question just wasn't clear to me. Basically, normally what I would look at is what is the difference at baseline to see if there was inadvertent confounding arising by chance. It looked like, if anything, there was bias against the BiDil group because they had higher blood pressure. So, the essence of the question is what was it during treatment where BiDil had reduced the blood pressure? Could we have achieved the same effect with another antihypertensive?

DR. NISSEN: I am going to answer that very specifically. One of the things we saw very clearly yesterday was, in the heart failure endpoint particularly, that there was a dissociation, some evidence of some dissociation between how you lowered blood pressure and the incidence of heart failure. We saw, for example, that if we lower it with ACE inhibitors you seem to

do better than if you lower it with calcium channel blockers on that particular endpoint. So, we have some evidence that suggests that you can't predict the effect on heart failure of a blood pressure lowering drug without testing it.

DR. FLEMING: Yes, that is a good point. At least the evidence yesterday indicated that the least clear association for blood pressure was with the occurrence of heart failure.

DR. NISSEN: That is right, as measured. It is an interesting question as to how much of this is antihypertensive, and I just don't think we know. We know that that can be dissociated at times in some drug classes.

Population--we are making some progress here. We might actually get dinner tonight!

DR. HIATT: On the second part of the question, I think the labeling should include the fact that it reduces blood pressure because of the concomitant medication issue.

DR. NISSEN: A-HeFT enrolled only the subgroup in which BiDil appeared to work in V-HeFT

I and II, namely, self-identified African Americans. How strong is the evidence that BiDil does not work in patients excluded from A-HeFT? If it were approved, what should labeling say about excluded subgroups; the underlying genetic or cultural bases for the observed differences? Let's stop right there for the moment.

This was the subject of much discussion in our open public hearing and I think we do have to weigh in on this. So, I will ask for the committee to offer some comments.

DR. PORTMAN: May I?

DR. NISSEN: Sure, please, courageously!

DR. PORTMAN: Well, courageously because, unfortunately, I have to leave. So, I am going to opine--this is the word of the day--on this particular issue as well as the voting on question 4 because I think that while I am doing an early termination of my meeting here, I have had a statistically significant amount of discussion to 0.00625! So, I apologize for leaving early but I agree with many of our public speakers that the

black population in this country is heterogeneous but I do believe, as a clinician, that there are certainly differences. I see it in my everyday practice. I see it in the African American kids who have focal sclerosis and who have hypertension and who have proteinuria. I don't know whether those differences are genetic or whether they are social or whether they are economic or whether health delivery-related and I don't think that particular issue is germane here.

But I do applaud the FDA for requesting the study be done in this population, as I do their request for studying 40-60 percent of pediatric patients in our antihypertensive trials as African Americans. And I do believe that the evidence that I have heard is significant enough that I think this drug should be approved.

I don't find any real justification for approving it only in one particular patient population. Granted, they haven't done the study in the white population but I think it ought to be approved in general, and I think that maybe we

should ask for a postapproval marketing study to look at the white population. But I think that the labeling should say that the study in an African American population. Then we can go further from there but that is my opinion.

DR. NISSEN: I am going to disagree with you, Ron, and I am going to try to articulate why. I recognize the passion and emotion that we heard from the microphone and I respect both points of view here. My view, by the way, is that drugs are not racist; people are racist and I think it is important to understand that.

You know, what do we try to do when we sit here at this committee? We try to identify whether a drug is efficacy and the population which the drug will benefit. We look at evidence. When the overwhelming evidence that leads me to believe that this drug is effective comes from a population that we can define by some characteristic--now, it is self-identified race in this case. That is very unusual. It is precedent setting. But it is the case. And we are moving forward in medicine toward

the era of genomic-based medicine. There is no question that in 10 or 15 years it is going to happen. We are going to have the ability to look for a snip that tells us that this group of people will benefit from X drug or will be harmed by Y drug. I know it has been predicted for a long time and hasn't happened yet but it is going to happen, trust me.

So, what we are doing here is we are using the background, that being from the African Continent and immigrating to the United States, although often not voluntarily; that the population that comes from that ethnic background seems to have some differences. We already know that there are differences. We already know that. We know, for example, that ACE inhibitors don't seem to work as well in African Americans. We know that certain diseases are more prevalent or less prevalent. So, what we are doing is we are using self-identified race as a surrogate for genomic-based medicine and I don't think that is unreasonable. I wish we had the gene chip. I wish we could do it on a genetic

basis. But, in the absence of that, we have some information that suggests that African Americans--we know that African Americans, self-identified, get a pretty robust response to the drug. Go ahead, Bob.

DR. TEMPLE: I can't get anybody to address the quality of this evidence but I want to raise it again. If we had no information about the white population and they just decided to work in the black population you could argue it is an under-served population and maybe there would be a good case for that. But we are not completely bereft of data about the white population. V-HeFT I and II is not such shabby data. It really makes it look, if you look at it collectively, like the response to BiDil is quite different in the two groups--quite different. Whether it is absolute zero in the white population, I don't know that but it certainly doesn't look like it is anything close to what we saw in A-HeFT.

To me, that is a relatively important point because a number of people have suggested

that, oh well, approve it for the white population. It probably works. We study drugs in mostly whites and we approve them for everybody. I don't think that is the present case. There is some data now. How strong you think the data is, is what I would like to her a little about. But I thought V-HeFT I and II are moderately convincing that this is not a very promising drug in the general white population. Could there be people in the white population? Probably but I don't know how to find them yet.

DR. OTA WANG: I would like to give another perspective.

DR. NISSEN: Please do.

DR. OTA WANG: Since you brought up genomics, in the sense that if we are going towards the movement of genomic sciences, I think that we need to really carefully look at the issue of self-identified racial categories because if the assumption is that these population differences are biological, the self-identified population is a social and political construct. If we are going to

say that genomics is on its way, if we are going to be looking for example the snips you talked about, that is going to be informed on a more biological basis and will be consistent with the assumption they are making about the population differences.

That would make me a little more comfortable than just using the self-reference. Because what I am hearing is that we are using the self-identity as a surrogate for a biological process. So, we may be satisfied with that, which I am not, but when it goes into the clinical situation that reference is going to be assumed by clinicians whether they ask the particular patient or not, and I think that inconsistency gives this a false notion that race has a biological basis on the data of the study that really isn't supported.

DR. NISSEN: I am jut going to comment here a little bit. I am going to comment a little bit here and say that we have already taken this step. Do you remember the discussion of the Life trial? We saw completely different results comparing atenolol with losartan, a drug that works

on the renin angiotensin system in blacks where there was a huge hazard in the group that got randomized to losartan in the black population and it was the other way in the white population. Why is that? For reasons that I can't explain there appears to be less susceptibility of African Americans, blacks, to the effects of renin angiotensin system drugs.

I reviewed at lunch with Tom Fleming the results of ALLHAT, and there was some pretty strong evidence in ALLHAT that there were differences in responses to this classification of drugs. So, yes, the road to hell is paved with biological plausibility but there is a biologically plausible explanation.

I am not as persuaded as Bob is by the evidence from V-HeFT I and V-HeFT II. These are trials done a long time ago without the kind of elegance in the sample size, and there are lots of issues there around that. But it certainly suggests that there are some differences. So, we have seen it before in ALLHAT. We saw it in Life.

We saw it in V-HeFT I and V-HeFT II. You know, I am not uncomfortable with that, knowing that we aren't there technologically. I wish we had the genetic markers to be able to use to decide who is going to respond to what drug but, in the absence of that, we have to use the best available evidence that is available to us today, and that evidence was used in this trial and it worked.

DR. OTA WANG: I guess I would counter. Is really the self-identification the best proxy for a biological process? I am not so sure that is the best proxy.

DR. TEMPLE: What are the choices one has? That is how people in V-HeFT I and V-HeFT II and A-HeFT identified themselves. All the analyses that showed differences in V-HeFT I are based on, for better or worse, that marker, as is also true in ALLHAT and everywhere else. No one has ever tried to do genetically, and probably would encounter difficulties if they did because they wouldn't know how. This is sort of what you work with because it is the best you have and everyone

agrees it is inadequate.

DR. OTA WANG: I guess I am going to go back to sort of challenging the committee then when we start considering the standards that we are going to use for scientific integrity of the study and the data before us, how much are we actually going to weigh that, given the benefits and consequences of what we are actually going to be deciding today.

DR. FLEMING: I guess I would have used exactly the term, Dr. Wang, that you used. I would call it a surrogate that we are, in essence, trying to target in an enriched population that there is some reason to believe would be most likely to have the most enhanced benefit to risk. I don't know that it is precedent setting to say that we will use parameters that are more readily measurable--age and stage of disease and race and other factors that, in fact, might not precisely characterize exactly what are the mechanisms or factors that lead to this intervention having a more favorable benefit to risk, but they are

practical and achievable and it is what we would typically do.

So, I hear what you are saying and it is relevant but I don't know that it is novel to say that we define as best we can, according to the measures that are readily achievable, who we think or what we think distinguishes statistically what are effect modifiers.

My answer to this basically is we have heard a lot of discussion about the fact that when we do trials there may be under-represented populations and, yet, we often will extrapolate those conclusions to as broad a group as with think that extrapolation could be justified. I distinguish that though from where we proactively exclude a subpopulation. When speaking with sponsors in designing trials I start with, well, what is your objective? What do you think the target population ought to be for this assessment, for this intervention? If you want to market this product in a given target population, then the study should be designed for that specific target

population. So, I can only presume that the sponsor here must have had in mind that the target population were those people who identified themselves as black because they proactively sought out to exclude everybody else.

Now, was there a rationale for doing that? I mean, obviously there could be a big debate about that, but I think that comes back to what Bob Temple was saying, and that is where I think the V-HeFT I and V-HeFT II data are really key. The primary analysis of V-HeFT II--I am going to follow the way the logic was that I understood from the sponsor, the primary analysis was the entire trial and that entire trial showed a relatively unfavorable effect against enalapril, a relative risk of 1.23 so a 23 percent higher rate of mortality over five years on BiDil.

But then there was a subgroup, probably many--race one of the key subgroup analyses, and there was some considerable evidence--not incredible but some considerable evidence of effect modification, relative risk of 1.01 in

blacks, 1.39 in non-blacks. What does that mean? It means that in the context of the population and the ancillary care that was delivered at this time there is a suggestion, at least, that if you are white you would be better off taking enalapril than BiDil; if you are black there is a suggestion that they are the same, that you would probably do just as well with one or the other.

Now, that interaction could occur either because there is effect modification for ACE inhibitors, or it could be because there is effect modification for BiDil, or both. There is not a lot of additional data on whether or on the effect of BiDil is specific to race. There is more data, and Steve was talking about some of those key sources for the class if we lump into the class, and yesterday we said it is aggressive to lump ARBs and ACE inhibitors but, if we do, then you look at the Life trial or you look at ALLHAT. I mean, ALLHAT had 8600 blacks and 15,400 non-blacks for the comparison of lasinopril to the diuretic and there was some pretty striking evidence that race

mattered.

So, I come away with that thinking that the data seem to suggest that there is particular benefit from enalapril in the white population and not no benefit but less benefit and BiDil is probably comparable. So, now we go to V-HeFT I. In V-HeFT I there is--with a placebo control so we now can see what BiDil's effect actually is--a trend toward an improvement in survival where that trend is somewhat more apparent in the blacks, although it is treacherous; it is based on 49 black patients treated with BiDil, 49 people. Okay? But there is that trend.

So, I am sitting where the sponsor is sitting. What do you have? What you have are two sources of information, V-HeFT I in the non-blacks, suggesting no difference between BiDil and placebo. Then, in V-HeFT II against ACE inhibitors it is not as good. If you believe in the interaction, it is not as good. So, that at least was the logic that I assumed the sponsor was following when they proactively decided not to test BiDil in whites.

What I really want to know and what I really like about A-HeFT is that they are answering the question about BiDil's role in the context of today's interventions which V-HeFT, not to its criticism but earlier point in time, didn't do. It is saying in the context of ACE inhibitors, and this to me greatly strengthens their results--in the context of ACE inhibitors, does BiDil add to that? Because if you didn't say that, I would have come aware from V-HeFT II as saying V-HeFT II says in whites use ACE inhibitors; in blacks ACE inhibitors are jut as good as BiDil so you don't need BiDil. But they answered the question if you give ACE inhibitors, because 75 percent of the people had that in A-HeFT, do you improve? If you believe the results are conclusive, what you are saying is, yes, they do improve in blacks.

But they proactively excluded the whites. What is the scientific basis to argue that we know that BiDil adds to ACE inhibitors in a white population? And I think this is where you are coming from, Bob--the data in V-HeFT I and II

certainly doesn't suggest that, although it wouldn't strike me as being irrational to test it. But I don't care if the issue is race or the issue is gender or the issue is age, or anything else. The principle that I am generally following is if you proactively exclude a population you are telling me in advance you expect that population to be less likely to have favorable benefit to risk. Then, if you study it in a targeted population, that is what your label should be.

DR. NISSEN: That is not what we do when we study drugs generally. That is not what we do because we say everybody can be involved, and we may only have 5 percent or 10 percent African Americans but we didn't exclude them. We didn't exclude them because we thought it wasn't going to work in those. This was a very specific experiment that was done and we have to be true to the scientific experiment that was done. By the way, it is actually more than you suggested. Over 90 percent of the patients got an ACE or an ARB. A lot of them got beta-blockers, diuretics. These

folks got the best therapy for heart failure available today. Then the experiment asked the question does adding this fixed dose drug on top of the best we can give patients today, does it do anything? My conclusion is that it does and some of the effects are pretty important, like on mortality. So, that has a lot of weight when I look at this. If these patients had been under-treated with conventional therapy, this would not be a very strong argument. I would have said why don't you just give them more ACE inhibitor? Why don't you treat more of them with ACE inhibitors? They did everything I would expect them to do in an absolutely state-of-the-art trial and that is a very compelling argument here for this population, but only for this population.

DR. TEMPLE: Tom said almost everything I can imagine. I just want to make one point. In V-HeFT II the black subset is ambiguous. You don't know if they don't respond to ACE inhibitors or not. But in the white subset is not ambiguous at all because we know that in general the white

population can respond to ACE inhibitors, but you saw when you broke them out that they had--well, you don't know that they had no response but they were 40 percent worse, which is as close as you can imagine to no response. I thought that was moderately strong evidence that it didn't look like it worked in that population.

That is somewhat important to me because I am worried about how much data we ask for in the population that isn't the one of major interest. So, I think that is an important principle that is coming up with more and more selection and individualization so I am worried about that. It is relatively important I think for us to conclude that there is at least reasonable evidence in that direction.

DR. NISSEN: One more thing, you know, we talk about this being precedent setting, it is not as precedent setting as you might think in that what we put in the label for losartan, based upon the Life trial, was that being randomized to losartan was a hazard if you were African American

compared to getting atenolol. You put that in the label, didn't you, Bob, as we suggested?

DR. TEMPLE: We did. There is another interesting phenomenon--I am sorry to get to sort of generalized regulatory stuff--in 1998 we required everybody submitting an application to us to do an analysis that took a look at effectiveness, safety and dose response in demographic subsets, generally meaning age, race and sex. So, we won't accept an application if it doesn't do those things. Pointedly, the intent is if something informative comes along you are going to use it.

There is actually a law--people connected with NIH can describe it better than I can--that requires NIH studies to gain information about demographic subsets, and if there is any suspicion that there might be a demographic difference they are supposed to design the study so it is large enough to actually tell the difference.

You know, for better or worse, I am sure there is a more sophisticated way to identify the

people of interest, or there will be some day, but at the moment there is legislation and practice and regulation that says until you can do better you are supposed to use this surrogate, bad as it is.

DR. OTA WANG: Well, actually, in a sense the logic is reversed. The assumption is that you are supposed to include everybody unless you can actually have a scientific justification to exclude people. So, it is sort of a reverse.

DR. TEMPLE: Well, you are also supposed to specifically find out whether there are differences in those subsets. That is what our regulation asks you to look for and labeling now regularly says we looked; we didn't find. Or, on rare occasions, we looked and here is what we found.

DR. NISSEN: I am going to take us back to the questions again. I think it was very important that we have that discussion because obviously this is a unique situation.

Bearing in mind experience in V-HeFT I and II, to what NYHA classes do the benefits of BiDil

apply? John, do you want to say anything about that? What classes should this apply to?

DR. TEERLINK: Well, I actually didn't answer the previous question which relates to my answer to this question--

DR. NISSEN: Please, do.

DR. TEERLINK: --as to how I view V-HeFT I and II in terms of their impact on this trial. I have been a bit torn because with V-HeFT I you have basically a negative trial. Then you go and you look at the interaction effect and you get a negative interaction effect for race. Then you say, okay, we will still keep looking and you drill past the interaction effect and then you find something that finds a beneficial effect. So, I am extraordinarily reticent to have that influence my decision too much, certainly in terms of a regulatory standpoint making claims on efficacy on the basis of those trials.

I know this is very challenging for public members to kind of comprehend. How can you have this positive result that you then don't take into

account? But we have seen multiple play of chance type issues, and I think we really need to protect against these play of chance issues when we are dealing with relatively small numbers of patients and obviously small event sizes. That applies to V-HeFT II, although that is a more complicated issue because of most of the issues that Tom said so I won't apply it to that.

So, then I am pretty much left with A-HeFT. I think your decision in terms of which NYHA class, which patients should be treated should be based on who were the patients that were enrolled in A-HeFT. So, that would be mu answer in terms of the NYHA class, which were mostly NYHA III.

DR. NISSEN: I agree with that. We have said many times how treacherous non-prespecified post hoc analyses are, even prespecified subgroup analyses are very dangerous, particularly in a trial that failed to meet statistical significance on its primary endpoint. That is the riskiest proposition of all. So, I believe it was Class

III/Class IV, was it not, John?

DR. TEMPLE: No, for practical purposes, it is Class III.

DR. NISSEN: Fair enough.

DR. TEMPLE: I just want to be sure, because we are going to have to deal with this, given the benefit in Class III and given what most people assume is a certain continuity of response, you don't think the results in A-HeFT, plus the assumption that things are not usually this continuous, plus the fact that V-HeFT I and II both had Class II and well as III, makes the case for identifying Class II and III patients as the target population? That is the sole issue, is it II and III or just III?

DR. TEERLINK: Just III.

DR. FLEMING: Tom, II and III or just III?

DR. FLEMING: Well, we are kind of blending in question 4 in now too. Do you want us to answer it only from V-HeFT or can I bring in--

DR. TEMPLE: You can do 4 first if you want.

DR. FLEMING: Basically, V-HeFT II was a II/III population and the Phase III trial is more advanced patients. The only comment I wanted to make here is just to point out that this is a surrogate too. This discussion here is, in fact, not fundamentally different from saying should the population be self-identified blacks versus not; should the population be NYHA Class III versus something else? Because it is trying our best to characterize a group homogeneous relative to benefit to risk, driven by certain risk factors that probably aren't so naively or simply characterized by whether you are New York Association Class III.

DR. NISSEN: I am sure you know this, guys, but New York Heart Association classification is pretty soft. It is continuous--what we are going to do is give guidance to clinicians about the population for which we know there are benefits. I am not sure even how accurately in clinical practice people actually classify. Patients come into my office and I don't write down

on a piece of paper what I think their New York Heart Association classification is. I sort of get a sense for it, and that is how people practice.

DR. TEMPLE: The package insert will describe the entry populations of V-HeFT I, V-HeFT II and A-HeFT if we approve it. So, that is not in doubt. The only question here is should the indications section, again if it is approved, direct treatment solely at people with Class III or something broader.

DR. SACKNER-BERNSTEIN: The study protocol limited the enrollment in A-HeFT to IIIs and IVs, and there are very few IVs. So, to be consistent it would need to be approved for III, with all the caveats that have been mentioned about the way classifications are assigned.

DR. NISSEN: Is there consensus about that?

DR. HIATT: Yes.

DR. NISSEN: Anybody disagree? Can we made that a consensus recommendation?

DR. TEMPLE: I didn't understand the

caveat part. What do you mean? Understanding it is for Class III but Class III doesn't mean anything? We should say that?

[Laughter]

DR. SACKNER-BERNSTEIN: That was not my caveat.

DR. NISSEN: Our answer is Class III. If there is dissent, let's hear it. No dissent--oh, there is dissent?

DR. HIATT: No, no dissent on that, but I do hear linkages between the V-HeFT and the A-HeFT trial, and I think the point has clearly been made earlier that this is add-on therapy in A-HeFT to excellent contemporary background therapy, and that is not the case for V-HeFT. I don't think the linkages should be all that tight. I think it goes back to this question about excluded subgroups from V-HeFT. I just don't think we know because if you add this on to current therapy in another group of patients defined by some characteristic, I can't tell you it wouldn't work. So, I guess from a labeling education point of view, I would not

associate V-HeFT with A-HeFT so much.

DR. NISSEN: You know, we are saying we are not so persuaded by V-HeFT in any of our thinking process and you may or may not want to factor that into your decisions. Nobody here seems to be all that impressed that it tells us much.

DR. SACKNER-BERNSTEIN: We had previously said as a group that it is almost like a Phase II kind of data set, and generally you don't put Phase II trials in there.

DR. TEMPLE: Yes, that seems clear. I must say, we were at least somewhat mindful of the idea that we are not sure people always know exactly what the difference II and III is. There is a certain expectation that effects would not be discontinuous over that, and that the V-HeFT studies gave you at least some reason to think--I totally agree with everything you have said. You said it is a different population, etc.--some reason to think that you might want to limit it to the Phase III population. But I hear a fair consensus that it ought to be.

DR. NISSEN: All right, moving along here, anybody want to comment on any other population specific differences? Do you want to say something? Please.

DR. SACKNER-BERNSTEIN: I am concerned about the relative lack of study of women with heart failure, African American women. I think that what we are doing here is we are using a surrogate and we have to be mindful of biologics. The biologic differences that are much better defined in cardiovascular disease relate to gender than to self-described race. So, you have a study that had, as best I can tell, jut a little over 300 women randomized to the combination therapy or placebo who were followed for over six months. The point estimate of risk of death is very favorable. The confidence intervals don't cross unity. So, you know, on its own it would look to be statistically significant. But we are looking at a very small population where there really are biologic differences.

And if we are talking about trying to

identify subgroups that are surrogates for biology, which is what we are doing by self-proclaimed skin color, then it would be inappropriate for us to ignore a recognized biologic character. Once we do that, then we have to say what does the internal data say about the persuasiveness of the effect for women and are there external data that we can use to understand better what the efficacy and safety is because of the relatively short follow-up.

DR. NISSEN: You know, Jonathan, I am going to disagree with you here and say a couple of things. One is that there are more women in this study than we usually see in trials of this kind, a lot more in terms of percentages. Secondly, trials are generally not powered to look for effects separately in men and women. That certainly was not a hypothesis to be tested here. So, unless we saw striking evidence of heterogeneity, compelling evidence of heterogeneity I would consider it to be not a relevant consideration. But maybe somebody disagrees.

DR. TEMPLE: Well, as Jonathan said, you

found the contrary. You found evidence of non-heterogeneity but still only in 300 people.

DR. SACKNER-BERNSTEIN: Yes, 300 people followed that long. There were actually more women than that entered. But there has certainly been a number of cases where drugs with an overall data set in a particular indication of 1000 to 1500 people which were approved, later were found to be drugs that had unanticipated effects for which there were safety concerns that erupted. So, when I see a population of 300 followed for that long, it raises a concern in my mind that I would rather have it corroborated by another study or another set of data.

DR. NISSEN: What implications would you make in labeling?

DR. SACKNER-BERNSTEIN: I think that we should be considering that labeling is for black men.

DR. TEMPLE: Are you ready to do another outcome trial in women?

DR. NISSEN: No.

DR. TEMPLE: With these data?

DR. SACKNER-BERNSTEIN: I think there are times in the past where there have been small data sets that have looked very good. To me, it seems inconsistent to accept the idea that we are looking at a surrogate for biologic differences and then we are choosing not to pay as much attention to a recognized biologic difference for which there is no external data to corroborate. So, am I really saying that I would only--if this were all in my hands--only approve this drug for black men? No, but I think that is something that is a very important issue not to ignore if we are going to be discussing these surrogates for biology.

DR. NISSEN: Let's see if there is any support around the table for anything in labeling about gender. I am afraid you are on your own on this one, Jonathan.

DR. TEERLINK: Actually, using Jonathan's logic--I don't know, are you using V-HeFT I to support the men? You must be using V-HeFT I because otherwise the boat you are in is you are

basically only relying on A-HeFT which has only a couple hundred men and only a couple hundred women. So, you are saying the drug shouldn't be approved at all.

DR. SACKNER-BERNSTEIN: Actually, what I was going to do before I voted on number 4 is ask Tom the question once again about the p values for mortality that he would consider to be the range within which they would fall in order to get a sense of how persuasive that is. Because I think it gets back to the issue raised about lupus and those other things. If you have a mortality effect that is compelling for a life-threatening illness, that is sort of the trump card. But it still is important, I think, to bring this kind of issue about true biologics into the discussion.

DR. NISSEN: Tom, do you want to say anything?

DR. FLEMING: Should I go ahead. Basically, going ahead is in the context of essentially answering question 4. DR. NISSEN: Why don't we do that? I think in the context of that

we can answer the question.

DR. FLEMING: Because I think it does come back to Jonathan's point.

DR. NISSEN: On this one we really need a formal vote I believe. So, please vote. We are asking now should it be approved. Then you can provide any supporting information you wish that might be helpful to Jonathan. So, we are going to go this way around the table.

DR. FLEMING: I guess my philosophy on all of this is what is valuable, hopefully, is the reasoning behind how we would respond to these questions as opposed to the vote. I would rather advisory committees didn't vote but just gave reasoning. So, let me give the most important, which is the reasoning, and the least important, which is the vote.

Essentially, my view--and I won't repeat it, but based on the logic to the answer to question 3, my sense is the V-HeFT I and II data, technically speaking, if you look at V-HeFT II, it is a trial suggesting an unfavorable result. When

you look within the subgroup and you find the neutral results in the black population and then when you go to V-HeFT I and look within that subgroup and find a favorable trend, to my way of thinking, all of that provided a very logical basis for the sponsor proceeding with the A-HeFT trial, which I think they then designed, and certainly in the context of assessing BiDil in today's state-of-the-art interventions was a very appropriate design.

But, essentially, where it leaves me is the approval for this in terms of benefit to risk in essence lies in how persuasive we view the A-HeFT trial to be. So, I look at that as something guided by the FDA and European authorities' sense of what it takes for a single trial. It is a study that is very well done, internally consistent, very persuasive statistical strength of evidence on the primary endpoint greater than just the 0.25 one-sided, and when I look at it in that context the primary endpoint is certainly a favorable result. Is it highly

compelling? Well, I am at least willing to say it meets strength of evidence of a single trial.

I have some trouble interpreting it, but I am greatly reassured when I look within the individual components. When I look at the mortality component my sense, from what we were saying before, is that the mortality data essentially is right at the strength of evidence of a single trial on mortality. It is right in that region of 0.025 one-sided, my interpretation of mortality.

For the CHF hospitalizations, to me, this is where the most clear positive signal is. In essence, what we are looking at was 85 against 130 patients that had a CHF hospitalization.

Somewhat dampening my enthusiasm about that though was looking at all-cause hospitalization because essentially, if I am following it, that 45-person excess of events in the control arm was cut to less than half, to something like 19 or 20 excess. So, there were 26 more people who then had a first hospitalization in

the BiDil group compared to the control.

However, counterbalancing that to an extent was the analysis, as put forward by the sponsor, that when you looked at time in the hospital, that was actually encouraging on all-cause hospitalization to the point that I think their p value was 0.01 for hospitalization days.

So, my sense is CHF hospitalization is the clearest signal. When you look at overall hospitalization it is disappointingly dampened. I want it to be the same. I don't necessarily expect it to be more; I want it to be the same. But I was somewhat reassured by days in the hospital.

In quality of life, I found that difficult to interpret, as I often do. But I consider it important information nevertheless because it is trying to get at the fundamental issue, and it is not just how long you survive; it is how well you survive.

Safety, while not totally pristine, is in my view relatively favorable. In the context of saying this isn't an analgesic; this is an

intervention that we are, in fact, trying to impact overall duration of survival as well as quality of survival.

So, when I look at the totality of all of these, in essence, there is some considerable consistency here. I would like to kind of conclude that there is, in a sense, something that is comparable to the strength of evidence of two trials where I get part of that out of mortality and I get part of that out of quality of life, and they are semi-orthogonal, i.e., it is semi-doubling or it is reinforcing each other. We are prolonging survival and we are improving quality of life among those people who have survived.

So, when I put the totality of that together, to my way of thinking it is a close call. It is a close call but in my view it does meet the general fundamental principle. What I wouldn't accept, what I wouldn't support is a label for having improved mortality; a label for having improved over hospitalization or a label for having improved overall quality of life. I would believe

that there one could justify a label for having improved heart failure hospitalization. That is what I think is most clearly established.

So, looking at the totality, while it is a close call in my view, I think the least important thing I have to say on all of this is I vote yes.

DR. NISSEN: Okay, Jonathan, you are up.

DR. SACKNER-BERNSTEIN: Can I ask Tom a question?

DR. NISSEN: Sure, but I want to move right through this; time is flying.

DR. SACKNER-BERNSTEIN: So, the p value that you would accept for mortality made it according to the idea of the single strength of the trial, and I know you have trouble with the composite but it seems as though you wouldn't want the drug labeled for those individual components.

DR. FLEMING: Well, it is certainly not a novel circumstance for us to be looking at a composite endpoint in heart failure. I can't remember the exact context but I do remember being asked by Ray Lipicky at one point, when we were

looking at heart failure hospitalization-free survival and we saw an effect of a composite endpoint, and it was very heavily driven by the heart failure hospitalization component rather than the mortality component, and what is the proper label. I think what he was driving from in that setting is the clear signal is what is heart failure hospitalization. To me, the clearest signal here is the heart failure hospitalization. The other components are reinforcing, hence, giving me a sense, even though there are some negatives--there is the negative of all hospitalization, but there is overall reinforcing. But I guess what I am arguing is to label it for mortality I would want more than marginal evidence on the strength of evidence of a single trial.

DR. NISSEN: Jonathan, your vote?

DR. SACKNER-BERNSTEIN: I have trouble as well with this composite and how to interpret it. But looking at the individual components, not only is the heart failure hospitalization persuasive and the days in the hospital for all heart failure

hospitalization, but actually I found data for all days in the hospital and there the gap actually widens. We were talking about how the gap in heart failure hospitalizations narrows. Total days in hospital for all hospitalizations seems to widen.

The quality of life, as it was presented in the graph that you mentioned, I think is very reassuring that drug really is having a favorable impact on people that is important and relevant.

The mortality benefit is one that at this level of significance I don't think you can ignore, no matter what your concerns are. So, the safety concerns that I would have can be watched for easily because we are talking about, you know, following blood samples in people who are affected.

So, I think that despite my reservations about the way the biologic surrogates versus actual biologic differences are handled in the population studied, the drug should be approved based on the data.

DR. NISSEN: John?

DR. TEERLINK: Hopefully, everybody has a

sense of how our real emphasis has been to protect public health by applying kind of a very rigorous and consistent approach to the data. My personal decision has been influenced by a long history of biologic plausibility and pathophysiology that we understand of heart failure. The previous drugs that we have understood and how they have worked within heart failure, the previous trials with nitrates and hydralazine in this context, and also the idea that we really want to make people feel better and live longer and that is our main goal.

In that context, I also think it is a very close call. It is a single trial having, for me personally, discounted V-HeFT I and V-HeFT II largely as an efficacy measure, we are left with A-HeFT which, on a mortality basis, I am convinced actually, because of the relatively small events, it is making it as a single trial, showing that there is a trend towards a positive trial in mortality, but it doesn't provide the usual kind of force of evidence so you would say, okay, this saves lives; this clearly saves lives.

However, there is a very consistent effect across the board and it is on the basis of the improvement in quality of life, importance in hospitalizations and a trend towards improving mortality that I would also recommend that it be approved. I think my recommendations in terms of how to label it would include self-identified blacks for now. I don't think that precludes that it will be useful in other populations in the future perhaps. I would also strongly emphasize that this trial was done in the context of ACE inhibitors and beta-blockers. I would want to make sure that there is no effort to say that this is an African American or black specific drug and, therefore, it should be the drug that is given to African Americans. I would actually think there should be direct wording in the labeling saying on top of standard therapy, especially ACE inhibitors/ARBs and beta-blockers. So, I vote yes.

DR. NISSEN: You know, I like these discussions because I learn a lot. I always learn a lot about how to interpret clinical trials, p

values and so on. But I think that there are times when one has to adjust one's thinking for the clinical factors and some aspects of this that I think are very, very important.

Let me see if I can outline them. First of all, I think this was a courageous thing to do, to try to develop a drug for this population which seems to have a disproportionate burden of disease. By the way, I did not agree with the speaker who argued that there isn't a disproportionate burden. I am convinced that there is. That is important.

You have already recognized that in some of the approaches to orphan drugs. It is not an orphan drug but it is a population that is a smaller subset in the overall population, and I do think it is more challenging when you exclude the majority of the population from a trial to get a big trial done. As a consequence of that, there are some issues about how much power the trial had. This was a trial that was on the margin and the power was further reduced by it being terminated early. I think the reasons have been discussed but

that does, in fact, hurt us in our ability to interpret the results.

Let me take one of the things that is most compelling for me. How high was the bar set? This trial set the bar about as high as you can set it. They took and they used beta-blockers; they used ACE inhibitors or ARBs and at rates that are almost unprecedented in other trials. They then asked the question. I am telling you, because I do active control trials, when you give the best therapy that is available today to a population and then you test on top of that, you are putting yourself through a very, very rigorous test. And that has to be considered in weighing the statistical evidence. You are not going to get p values of 0.0001 in a non-placebo controlled sort of an environment where you are not giving any other therapies. It is hard. It is hard unless the therapy is incredibly powerful.

Now, it turns out that the therapy was powerful because the point estimates here show very large benefits. I said earlier that if there was a

15 percent reduction in mortality it would to be a home run. This was 43 percent. I know the confidence intervals are wider than we would like, but that point estimate does tell us something, and it is certainly very useful.

So, I have to approve the drug when I think that there is evidence that we could reduce mortality by as much as 43 percent, and if you use the upper confidence intervals it could be even more than that in a population that is already on the best therapy that is available today for congestive heart failure.

So, I am not as conflicted as several of you are. I find the evidence more compelling. Compelling doesn't necessarily mean statistical. Compelling to me means also clinical and, as a clinician, I find the evidence more than adequate to vote for approval. Thank you.

MR. SAMUELS: Thank you for the opportunity to participate in what I consider to be a historic decision. I think I said earlier about the fact that, you know, we are talking about a

population that is over-represented in terms of disease and under-represented in terms of being able to have things to fight those various diseases.

I, indeed, am impressed by the mortality statistics. I definitely vote in favor. However, I would urge that the labeling be identified as having clinical trials for people who have identified themselves as being African American.

DR. NISSEN: Thank you. Bill?

DR. HIATT: I too think it is approvable. I don't want to reiterate a lot of the comments that have already been made, but I think there are two lessons I take from this that are worthy of discussion. The first is that given the compelling evidence from the V-HeFT studies, the sponsor was given a direction from FDA to do a single study to meet approvable criteria. We discussed what we think that means.

In that context, I think the study seems to have been under-powered. Certainly during the early looks it was under-powered and I think,

Steve, you mentioned that as well. I guess I would say to future people going forward don't do that. It may be resource limitations but the power issue was a problem for me.

The second was the conduct of the DSMB. I discussed that earlier, but I think if you are going to be looking at the data and you might be starting off under-powered, you should be very careful how you then do the multiple looks that Tom has spoken to. Ultimately then there are questions about how convincing the data are. In the end, I agree clinically they are convincing. I don't think we need more data right now. But I think in retrospect there are some conduct issues that could have been avoided, which would have made this a little bit cleaner in terms of our discussion today. But I would vote to approve.

DR. NISSEN: Susanna?

DR. CUNNINGHAM: Thank you. Well, as you know, it is my challenge to represent the consumer and, as I see it, the consumer doesn't have just one voice. The consumer has multiple positions.

There are at least three positions, and there are probably more. One is, please approve as indicated or as proposed. Another one is to please approve but not restrict it to any population. The third one is please do not approve.

Those are the possibilities through the eyes of the consumer, and some of the issues that consumers have, and I think maybe one of the most important ones that we can't answer here but we need to continue to do all we can to address disparity in healthcare. That is really a key underlying concern.

I also think the point of giving

"illusion of inclusion" is something we really need to think about. Usually I am here asking for representation of ethnic groups and gender, and I think in this case we do have data and that is unusual and refreshing in some ways.

There is also the issue of just having one trial, and it can be seen as finally attending to the needs of this population. It could also be not

having a bar that is high enough. So, I think we have the possibility of not being satisfactory whichever way we go and I think that is a problem we face.

There is the problem of how adequate race is as an indicator. We have had a lot of discussion about that. I think an important thing is the issue of quality of life and the improvement of survival. Please keep me out of the hospital, you know, being able to breathe and sleep at night, not having to use three cushions or three pillows is pretty important.

I think the issue of labeling for blacks when we never label for Caucasians actually has been answered somewhat in the discussion of populations and how this population was particularly selected.

And I think the need for effective treatments, as Mr. Samuels has said so eloquently, I think that is very important and for that reason I would vote yes.

DR. NISSEN: Thank you. Dr. Ota Wang?

DR. OTA WANG: Overall, I think that my decision is really based on a combination of the pretty intensive statistical discussions we have had about the data and trying to understand that within the clinical context. So, my decision is really based on a combination. I think some of the design issues and the statistical analyses really--on that basis alone I would feel very uncomfortable to actually approve. But I think the qualitative discussions around the quality of life are very persuasive. So, on that basis I would actually approve.

What I do not agree with, with everyone who has spoken so far, is having a label targeted towards a particular group that I feel has not been--I don't find the evidence persuasive. I think that we should ask the sponsor specifically to discuss at least the reasoning about their biological understanding of their self-designated racial reference is important because the underlying rationale for the A-HeFT is, in fact, a biological reason. And, because of that, I think

it will be over-interpreted, whether we have that intent or not, as a biologically related drug. Because of that, I do not support having a specific label targeting the African American or black people.

DR. NISSEN: That completes the vote. Before I adjourn us, let me say this is my last meeting. I retire from this committee on July 1, and I really want to thank all of you for the privilege of serving with you on the panel. I also want to particularly thank the FDA. I have served through three division directors. I have learned a lot from all of you and I appreciate how hard you work to make these committee meetings meaningful and useful and helpful. I have learned a great deal in doing this and I do appreciate the opportunity to have done this. So, thank you very much.

DR. TEMPLE; We want to thank you and the entire committee for two very interesting days. I am sure we are going to follow-up with some projects internally and we will get back to you,

but a lot of interesting issues have been raised today and yesterday.

DR. NISSEN: Thank you.

DR. TEMPLE: Thanks.

[Whereupon at 4:46 p.m., the proceedings were adjourned.]

- - -