DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

PERIPHERAL AND CENTRAL NERVOUS SYSTEM

ADVISORY COMMITTEE

ISSUE: SAFETY AND EFFICACY OF FREEDOX

TIRILAZAD MESYLATE INJECTION (NDA 20-399)

Thursday, April 29, 1998

8:30 a.m.

Holiday Inn Gaithersburg
Two Montgomery Village
Gaithersburg, Maryland
Bethesda, Maryland

MILLER REPORTING COMPANY, INC.
507 C Street, N.E.
Washington, D.C. 20002
(202) 546-6666

PARTICIPANTS

Sid Gilman, M.D., Chairperson
Sandra Titus, Ph.D., Executive Secretary

MEMBERS

Michael H. Brooke, M.D.
David A. Drachman, M.D.
James C. Grotta, M.D.
Claudia H. Kawas, M.D.

NEUROLOGICAL CONSULTANTS

Ella P. Lacey, Ph.D.
Richard D. Penn, M.D.
Gerald Van Belle, Ph.D.

FDA

Greg Burkhart, M.D.
Lu Cui, Ph.D.
Russell Katz, M.D.
Armondo Oliva, M.D.
Judith A. Racoosin, M.D., MPH
Robert Temple, M.D.

C O N T E N T S

P R O C E E D I N G S

Call to Order, Introductions

DR. GILMAN: I welcome you, all of you. My name is Sid Gilman. I am chairman of this committee and I am Professor and Chair of the Department of Neurology of the University of Michigan Medical Center.

Before we go around and introduce ourselves, let me just state a few ground rules concerning this meeting today. I ask members of this committee to raise your hand or signal in some way that you would like to ask a question so that I can recognize you, and we can have an orderly discussion.

For the agency, please allow us to interrupt your presentations with questions as we go along, and please do your best to answer our question directly at the time that it is asked instead of postponing it until later. Sometimes those questions get lost in the course of the conversation.

The sponsor has asked that it be allowed to make its thirty-five to forty minute presentation without interruption unless there are questions about their slides, and I said that we would do our best to accommodate that wish. That's fine. Since the FDA will present its overview first, I suspect we will be asking more questions of the FDA than we will of the sponsor but, again, if questions arise we will need to interrupt and have those questions

clarified.

With that, let's go around the table to introduce ourselves. We will start with Dr. Drachman.

DR. DRACHMAN: I am David Drachman, from UMASS Medical Center.

DR. KAWAS: Claudia Kawas, from Johns Hopkins.

DR. PENN: Richard Penn, from Rush in Chicago.

DR. LACEY: I am Ella Lacey, emerita faculty, Southern Illinois University, Carbondale, Illinois, consumer representative.

DR. VAN BELLE: Gerald Van Belle, from University of Washington in Seattle.

DR. GROTTA: James Grotta, from the University of Texas in Houston.

DR. TITUS: Sandy Titus, with the FDA's advisory committee staff, and I am executive secretary for this committee.

DR. BROOKE: Michael Brooke, Professor of Neurology at the University of Alberta in Canada. I spent fifteen years at Washington University in St. Louis.

DR. RACOOSIN: Judy Racoosin, medical officer, safety team, Division of Neuropharmacological Drug Products.

DR. OLIVA: Armando Oliva, medical officer, Division of Neuropharmacological Drug Products.

DR. BURKHART: Greg Burkhart, safety team leader

in the Division.

DR. KATZ: Russ Katz, Acting Director of the Division.

DR. GILMAN: Good. I will ask the sponsor to introduce yourselves. There are probably twenty-five or thirty of you out there.

All right, Dr. Titus has a conflict of interest statement to read.

Conflict of Interest Statement

DR. TITUS: Regarding the conflict of interest for Freedox, the following announcement addresses the issue of conflict of interest with regard to this meeting and is made a part of the record to preclude even the appearance of conflict at this meeting.

Based on the submitted agenda for the meeting and all financial interests reported by the participants, it has been determined that all interests in firms regulated by the Center for Drug Evaluation and Research which have been reported by the participants present no potential for a conflict of interest at this meeting.

We would, however, like to disclose that Dr. Claudia Kawas' employer, John Hopkins University School of Medicine, was previously involved in a study of Freedox. Dr. Kawas had no involvement whatsoever in the study.

In the event that the discussions include any

other products or firms not already on the agenda for which an FDA participant has a financial interest, the participants are aware of the need to exclude themselves from such involvements 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 involvement with any firm whose products they may wish to comment upon.

DR. GILMAN: Thank you. We will turn next to Russell Katz, the Acting Director of Neuropharmacological Drug Products.

FDA Introduction

DR. KATZ: Thanks, Dr. Gilman. Welcome back to the committee. I am glad you decided to come back.

My purpose here is just to give a very brief introduction to the issues that we would like the committee to consider when discussing this NDA. You will hear much more about the data in detail from both the agency and the sponsor. The questions that we have posed to you, the formal questions are very simple, but we know from yesterday that that can change. But, beyond that, they actually sort of mask many complexities in the data and issues that arise. So I just want to make those issues explicit. These are the ones that we have identified, at least up to now, and I just

want to run through them briefly.

As you undoubtedly know, the history of the application is somewhat interesting. It was originally submitted in 1994. It contained the results of two controlled trials, study 32 and 29. Study 32 did not show a statistically significant effect on its primary outcome, which was the incidence of cerebral vasospasm, but it did show an effect on mortality. When we looked closer at the data, it actually was the case that the entire effect seemed to be coming from men. There was no effect in women; no trends in women. It wasn't entirely clear at the time why that was so.

Based on this finding, the primary outcome for study 29, which was ongoing at that time, was changed from cerebral vasospasm to mortality, and that was actually negative as well but there was a trend in men in a retrospectively created subgroup. This was the subgroup of patients who were sickest at baseline, the so-called neurogrades IV and V.

The application at that point was brought to the committee and the agency ultimately, after the committee's discussion and its own review, sent a not approvable letter in 1995, I believe, which said that there seems to be a statistically significant finding on mortality in men but before we can ascribe this effect to drug treatment it needs

to be replicated, and you need to do a study in women to provide the substantial evidence of effectiveness that you need for approval. The effect on women that we were asking for was effect on mortality and favorable outcome. The letter didn't mention anything about subgroups. The intention was all women because the effect was seen in all men.

Nonetheless, as you know, subsequently the sponsor did two trials in women, study 65 which was intended to look at mortality as an outcome in women and only studied women. It studied women at a higher dose because the thought was one possible reason why women didn't show an effect in the earlier studies was that they were under-exposed, because of difference in metabolism, compared to men.

Study 65 was negative on its primary outcome, and on the basis of analyses that the sponsor did, they decided before unblinding study 63 to call the primary outcome mortality, as it had been, but only in a restricted subgroup of these neurograde IV and V patients, the sickest patients. In that study, when that outcome was amended prospectively, before the data were looked at, that had a statistically significant outcome.

So, there are a number of issues that are raised by this history and the data package, and the first has to do with the critical question of replication and whether or

not there is a bona fide finding that has been replicated.

And, there are two sort of related issues with regard to that replication, it seems to us. One is whether or not there has actually been a bona fide finding found in this subgroup, which is what the sponsor is asking us to approve the drug for. We say this because study 63 showed a prospective, of a sort, statistically significant p value but study 32 did not. It was positive in the high neurograde men in study 32 but it was also quite positive in the good patients, the neurogrades I through III. So, the question is raised as to whether or not there is a bona fide finding in this subgroup. Clearly, the effect was restricted to that subgroup in women in study 63 but it was clearly not restricted to this subgroup in men. That is the first question about replication.

The other question is sort of an interesting problem, and that is that even though in three out of the four studies that have been done the high neurograde patients were a retrospectively created subgroup, there are trends, numerical or statistically significant depending on the study, in that subgroup across the four studies. And, the question is whether or not these retrospective findings, all in the same subgroup, are equivalent, if you will, to the sorts of usual types of replications we like to see which is at least two trials which prospectively designate a

primary subgroup or primary outcome. So, the question is whether or not the same subgroup, retrospectively created three out of four times, provides the sort of substantial evidence that we ordinarily look for.

Another issue has to do with the identification of the subgroup to be treated. The law basically requires that labeling be able to written that adequately describes the population in whom the treatment is intended. The sponsor has produced a subgroup on the basis of a neurograding that is based on a somewhat idiosyncratic use of the Glasgow Coma Scale. Of course, you will hear much more about it and you have read about it. And, the question is whether or not patients can reliably be assigned to this subgroup so that labeling can adequately describe who needs to get this. This is particularly a question with regard to patients who were intubated, and I go into some detail in my memo in the file about how the assignment to neurograde in those patients is somewhat questionable, and I am sure we will hear more about that from the company but, nonetheless, that is an issue as to whether or we can adequately describe in whom this drug is intended to be used.

That is a generic concern. It is made more immediate because we think we have identified possibly a signaled increased risk in the complementary subgroup. In other words, in patients in I through III we feel that there

is a signal that there is an increased risk. So, the assignment -- the correct assignment to neurograde is a critical question when you think that patients who might be treated inappropriately might suffer some harm. So, we want to know what you think about that as well.

Theoretically one could write labeling. If you found that there was substantial evidence of effectiveness in the severe neurograde subgroup, it is theoretically possible that labeling could be written to describe -- and if you found that there was a risk in the lower subgroup -- you could write labeling to adequately warn, theoretically anyway, people not using it in this subgroup but using it in this subgroup because there is risk in one as opposed to the other.

I would just point out in that regard that there are a number of examples in other drugs where labeling, despite warnings and various other attempts to inform prescribers, doesn't necessarily preclude the use of a drug off-label. So, one thing we want to ask you about is, even if you found that there was substantial evidence of effectiveness in the neurograde IV and V patients, whether or not you think there is a risk in the I through III patients, and whether or not that risk is so severe as to preclude the approval of the product even, as I say, if you found there was substantial evidence in the high

neurogrades. So, that is another issue we would like you to discuss.

Another issue is sort of the ease and the practicability of actually assigning patients. As I say, one has to calculate a number on the GCS and translate it to a neurograde. It is perhaps particularly complicated when one also has to apply the Hess Scale to decide whether or not patients ought to get nimodipine which is, as you know, approved for the patients who are in a reasonably good condition after subarachnoid hemorrhaging.

So, the fact that a prescriber would have to calculate two scales and assign patients to various treatments is an issue that we would just like you to discuss as well. That raises, of course, the question of concomitant nimodipine. Even though it is approved for patients in neurogrades I through III, even though that is a different grading scale, there is a question as to whether or not -- well, first of all, we know that in these studies it was used in all neurograde patients, and presumably that is what happens but it certainly was true in these trial populations. So, it raises the question of concomitant use of nimodipine in the neurograde IV and V patients, which are the patients, of course, we are interested in here.

There is some evidence that in those patients nimodipine has a deleterious effect. There was a controlled

trial described in the labeling for nimodipine which suggests, I believe, a statistically significant increase in mortality in those severe patients and, in fact, the drug is not recommended for use in those patients. Now, that study was done at a higher dose than the approved dose for nimodipine but, nonetheless, there is this study. We have looked through the literature to see if there are other studies that speak to the question of whether or not there is a deleterious effect of nimodipine in high risk patients and we have not been able to find any other trial that speaks directly to this issue. So, there is this outstanding concern that nimodipine might make the high neurograde patients worse.

If that is true, then that has profound implications for the interpretation of the trials that we are presented with today, and that is something that we would definitely like you to discuss.

Next is the issue of differences at baseline between the treatment groups and how that affects the interpretation of the data. Now, both the sponsor and the agency have investigated these baseline differences for various reasons. Ordinarily, this is a practice that we are very suspect about. We ordinarily don't like to do this; we don't like sponsors to do this. So, ordinarily the context in which we see this is when a study is negative and

sponsors try to suggest that, well, the reason they are negative is because there were baseline differences in important prognostic factors, and they do analyses that adjust for these factors and, all of a sudden, something is positive.

Well, in study 63 the study is positive by the protocol in the high neurograde patients but we have looked at the baseline differences in important factors and, as I say, we ordinarily don't like to do that. The reason it was done in this case is because the randomization for that trial was not stratified by neurograde. You recall that that was a study in all neurograde patients and the high neurograde patients were chosen to be the primary subgroup of interest in a blinded, prospective way but basically at the very end of the trial. Everyone had been enrolled, I believe, and almost all patients had completed at that point. So, the randomization had not been stratified on neurograde; it wasn't an issue at baseline.

There is a view in the review team that extracting a subgroup from a trial that did not have a stratified randomization increases the likelihood that there will be important differences at baseline in certain factors between the treatment groups in that subgroup. There is also a view within the review team that it doesn't increase the likelihood. I mean, one view is that had you taken those 154

patients and randomized them from day one to treatment or control you would have the same chance of having baseline differences as if you had extracted them out but that is a controversial point, and we need you to discuss that. And, it is important because when you adjust for baseline differences in those neurogrades you can get analyses that eradicate the statistically significant difference that the sponsor found. So, that needs to be discussed I think in some detail.

As part of the effort to look at the because differences, another finding emerged. Our statisticians noted that whatever finding is there in the neurograde IV and V patients, it seems to be coming entirely from an even smaller subset of patients. These are patients that have been labeled PMR2 by the statistician and these are patients who have bilateral poor motor responses. So, the neurograde IV and V patients are a small subset but the effect, if it is there, seems to be coming from an even smaller subset. And, the implications of that finding for questions of replications and adequate labeling are interesting and we would also like to you to address those.

Finally, from the point of view of effectiveness issues regarding the evidence of effectiveness is the question of the integrated analysis that the sponsor did, and they did this for various reasons, one of which was to

sort of shore up the argument that there is evidence of effectiveness. Also, they attempted to identify any treatment by various factor interactions to see whether or not effects are coming from various subgroups, and subgrouping according to these factors had an effect on the outcome.

They did a similar sort of meta-analysis the first time around with studies 32 and 29 in the initial application. The committee and the agency didn't find it terribly helpful or contributory, and we would be interested to know what you think about it, particularly with regard to some findings which were negative on face which now the sponsor generates nominal p values for. For example in study 65 which was negative in the neurograde IV and V subgroup, I think they have identified a factor, initiation before or after surgery, which suggests that in neurograde IV and V in study 65 for patients who got the drug before surgery there was a nominally significant p value. So, we are very interested to know what you think of all of those analyses, all of which, of course, are retrospective.

Then, of course, there is the question of safety. As you know, development of the drug was halted in two indications, head trauma and stroke, at the time because of a signaled increased mortality, and our safety team has looked at that and we believe that there are replicated

findings of increased mortality in those studies. They suggest that there is some specific, although at the moment to us ill-defined, CNS toxicity, and a look at the subarachnoid hemorrhage experience suggests that there are also similar types of CNS toxicity. So, we are very interested to hear what you think about that.

In addition, of course, there is the potential increased toxicity in the neurogrades I, II and III patients, and also increased mortality seen in a dose-related fashion in study 7 and 19, which were two small early subarachnoid hemorrhage studies but which we also think showed a very interesting, as I say, dose-related, in one study statistically significant increase in mortality.

So, that is our least of issues we would like you to discuss. I am sure more will emerge, and with that I will turn it back to Dr. Gilman.

DR. GILMAN: Thank you. Before you leave the podium, Dr. Katz, one aspect of this history puzzles me a bit. So, in study 32 there were no differences between placebo and drug for the entire group, but retrospectively the men at severe grades showed a significance --

DR. KATZ: Well, all men showed a significant increase in mortality.

DR. GILMAN: Excuse me, yes, all men did. In study 29 though, even though the primary outcome was changed from

vasospasm to mortality at 3 months, the analysis at day 76 showed significant effect but at day 91 it did not show a significant effect.

DR. KATZ: Yes. I think actually it is day 106. The confusion arises around the fact that mortality at 3 months was the primary outcome --

DR. GILMAN: Right.

DR. KATZ: But 3 months was defined as plus/minus 2 weeks. You could have your 3-month visit at day 76 or you could have your 3-month visit at day 106. Now, if you looked at day 76 there was a nominally significant -- it is all retrospective at this point --

DR. GILMAN: Right.

DR. KATZ: -- but it was nominally significant in men if you used data from day 76, but we asked the sponsor to get additional follow-up data out through that entire 3-month window, which was actually a month from day 76 to day 106. If you look at day 106 there was an additional death in men. So, I think it was originally 1/20 and now it was 2/20 or something like that, and that nominal significance even was lost at that point.

DR. GILMAN: My question is having lost significance at that point with one retrospective study showing a benefit in men and the other one maybe yes at day 76 for grades IV and V but then that significance was lost

later on, why did you not request additional studies in men?

DR. KATZ: Well, I think we thought that the effect seen in men on mortality, at least in study 32, again not the primary outcome, was fairly robust. You have the numbers there in front of you. We just thought it was a very robust finding; not enough to attribute it to drug because it hadn't been replicated, but there are many reasons why studies of drugs that are even effective are not replicated. You know, studies are negative even for effectiveness treatments. So, we felt that 29, even though it was negative on its new outcome measure of mortality, it didn't really necessarily completely negate the finding in men in study 32, and we felt that if that finding could be replicated in women, all women, you would have one study with a robust finding in men, you would have one study, hopefully, with a robust finding in women and that would sort of provide you with the replication that you would need to be able to indicate the drug for all patients.

DR. GILMAN: Other questions from the panel? Dr. Van Belle?

DR. VAN BELLE: I just want to make sure, from you letter of April 14 I get the impression that the company was asked to provide evidence of efficacy in females, that the agency had considered the effectiveness in men to have been established by study 32.

DR. KATZ: Well, again, established is an interesting word. We felt that if that same finding were replicated in women that would provide substantial evidence of effectiveness. So, when we say established, if we thought it was established in the sense of drug related we would not have asked for any replications. So, we thought it was a statistically significant difference but that having shown it in only one trial it did not convince us that the effect was treatment related and that was why we asked for replication.

DR. VAN BELLE: Well, your statement here is that study 32 did demonstrate an effect on mortality in men only. I take that to mean a treatment effect.

DR. KATZ: A statistically significant difference -- let's say that ordinarily without replication we are hard-pressed to say it is treatment related, but the point is that if we thought that was a significant finding in men and if it were replicated, I suppose, a second time in men you would have had two trials in men and that would have convinced us that the drug was effectiveness in men. I think that is the point.

DR. GILMAN: Dr. Penn?

DR. PENN: I think I am one of the few people that was at that first panel meeting on this subject, and I remember taking away the opinion, at least of the panel,

that the drug had been shown effectiveness in men and that the major problem was having a drug that could only be used in men, and that the real question was women. So, the panel was not divided substantially about that particular issue.

The question I have for you is you used the words nominally significant throughout your presentation. Could you tell us what you mean by nominally significant?

DR. KATZ: What I mean by that is these are usually p values that are generated in subgroups or various outcomes of subgroups that were not prospectively designated so that you can generate a p value; you can test that; you can get a p value that is less than 0.05 but it doesn't really have the same interpretation. It may have no interpretation but it certainly does not have the same meaning as a p value less than 0.05 in a prospectively designated outcome or prospectively designated --

DR. PENN: So you might call it mathematically calculated --

DR. KATZ: As opposed to a p value that is really interpretable.

DR. GILMAN: Dr. Drachman?

DR. DRACHMAN: The rationale for women not being successfully treated was that there was a different rate of metabolism and different blood levels of drug, really based on young women tested during an initial phase. Why didn't

you request that blood levels be done during these trials rather than saying that a larger dose really was needed?

DR. KATZ: Why didn't we ask if blood levels were taken in which trials? The subsequent trials in women?

DR. DRACHMAN: Right, yes.

DR. KATZ: Well, I don't recall before those trials were done -- I am sure there are many people in the room whom do about whether or not plasma -- the studies were done at 15 mg/kg/day and the high dose in men had been 6. I am sure there was some work that looked at the comparability of plasma levels, and the dose of 15 was chosen presumably on the basis of the fact that it gave essentially equivalent plasma levels to a dose of 6 in men but, again, there are people here who know that data far better than I do. It was chosen for a reason. It is interesting that it is really only in premenopausal women that that difference occurs. Also, that difference even in those women, I believe, is essentially gone by day 4 or 5 of treatment. There are differences in the plasma level of the metabolite after that but the parent drug actually sort of equals out.

DR. GILMAN: Dr. Temple has joined us now, Robert Temple.

DR. TEMPLE: Just on the matter that Dr. Penn raised, we did not at the time we refused to approve the application think that effectiveness had been established in

men because there was really only one study that showed that. The first study was actually overall significant as well and, being suspicious of subset analyses, I think we would have said that was one favorable study. The oddity was that all effects appeared to be in men, who were only 30 percent of the population, which was weird.

But the second study didn't quite make it. It was a retrospective subset analysis and statistical significance at least was dependent on exactly what day you cut it, and the finding was in a very small subset of the population anyway. But we really didn't think it was established for any population. The reason to look in women was that was a group that hadn't been sufficiently studied and at the time that seemed a plausible -- I don't know how plausible it seems in retrospect -- explanation for why they might not have responded in the same way. So that would have been considered a confirmatory study for people. I mean, it would have been done in women but it would have shown the ability to replicate this finding in people with an appropriate adjustment of dose because it was thought that women metabolized and got rid of it and cleared it more rapidly. But we were not convinced that there was a finding in men. If we were, we would have labeled it for men.

DR. DRACHMAN: Again, that raises the question that I posed originally to Dr. Katz. Then why did you

recommend a study in women only? Why not in a larger population that included men?

DR. TEMPLE: That could be done but it was considered that a finding in women would constitute a replication and would have made the finding previously seen in men plausible -- stronger.

DR. DRACHMAN: Again, we are talking retrospectively but then you would have had one positive finding in men; another not positive; and then you would go on to women and you would wind up in a similar situation which, in fact, is close to where we are.

DR. TEMPLE: It depends a little on what you think of as not positive. The second subset analysis that turned on one more death -- I mean, it was 35/5 and 30/10 in a very small subset -- you could easily have decided that constituted additional evidence.

DR. KATZ: But I think at the time we were looking at just the two studies, 32 and 29. I don't think we put any stock at all in this neurogade IV and V subgroup at the time. Now looking back after four studies were done, this finding seems to be sort of emerging from this subgroup across trials and, obviously, we will talk about that and, of course, that is the claim that the sponsor is going for so it has taken on importance. At the time it was just something that had been identified retrospectively and was

really not taken very seriously.

DR. GILMAN: Any other questions for Dr. Katz or Dr. Temple? If not, we will move on to Dr. Oliva. We do have copies of these slides before us.

Efficacy Data

DR. OLIVA: Good morning. I am here to discuss the efficacy data. For those of you who have copies of my slides, I should point out that, with the benefits of modern technology, I had the luxury to make some last-minute corrections and changes last night. So, you will notice some, hopefully, minor differences in what is on the screen and what is on the page.

[Slide]

My talk is divided into three sections. First I am going to discuss important background information. I will breeze through this as Dr. Katz covered this already. Then I will discuss in some detail the efficacy results of the four large multicenter studies. Then I will conclude my talk with a discussion.

[Slide]

This slide shows the sponsor's proposed indication for tirilazad if approved. Tirilazad would be intended for the treatment of aneurysmal subarachnoid hemorrhage to improve survival and functional outcome in patients with poor neurologic function following the initial hemorrhage.

The draft product labeling recommends the initiation of treatment within 48 hours.

[Slide]

Just a little bit about the compound, tirilazad is a member of a new class of synthetic, non-hormonal 21-amino steroids which has been studied extensively as a cytoprotective agent and as an inhibitor of membrane lipid peroxidation against the damaging effects of CNS injury, including subarachnoid hemorrhage, ischemic stroke, closed head injury and spinal cord injury.

It is available as an intravenous formulation. It has a very long half-life of approximately three to five days, and it is highly protein bound. It undergoes primarily oxidative metabolism in the liver and is excreted in the bile. There is an increased clearance of the drug in premenopausal women by about 40 percent compared to either young or middle aged males. This results in lower drug exposures in these women. This gender difference is important when interpreting the results of the early efficacy studies, and I will refer to this point later. I should also point out that the gender difference is not as great in postmenopausal women, although we do see some, suggesting that it is more closely related to menopausal status.

[Slide]

The NDA, as you know, was submitted in June of '94. The original submission contained the results of two large, adequate and well-controlled trials that examined the efficacy of tirilazad in subarachnoid hemorrhage. These were studies 32 and 29, as you now know.

Study 32 was completed first. This study was negative on its primary outcome which was vasospasm. However, there was a positive effect on mortality seen with the highest dose, and this was in everybody. But on further analysis it did appear that the positive effect on overall mortality was actually a reflection of the drug effects on men, and there seemed to be no significant effect on mortality in women. Study 29 failed to reproduce this finding, and I will review these results later in a little bit more detail.

The Peripheral and Central Nervous System Advisory Committee met in September of '94 to discuss these results and, after much discussion the committee did not formally vote because sufficient evidence of efficacy had not been submitted.

The FDA issued a non-approvable letter in June of '95 in which the agency acknowledged evidence of a positive effect on mortality in men in study 32, but it stated that efficacy in women would also need to be demonstrated prior to approval, both in terms of improved mortality and

functional outcome. Additional safety data were also requested. The sponsor then conducted studies 65 and 63 only in women.

[Slide]

With that, I will proceed to discuss the efficacy studies. These were the four studies. Although the results of studies 32 and 29 were presented at the last advisory committee meeting, I believe they are pertinent to today's discussion and I would like to describe them briefly here as well.

[Slide]

All four studies showed a similar design. They were randomized, double-blind, vehicle-controlled, multicenter trials. Patients were enrolled with a diagnosis of aneurysmal subarachnoid hemorrhage due to a ruptured saccular aneurysm. The diagnosis was confirmed by angiography, and treatment was started within 48 hours of the bleed. All neurogrades were treated, and I will discuss the determination of the neurograde shortly. All patients received concomitant nimodipine either orally or intravenously.

[Slide]

Tirilazad or vehicle placebo was given intravenously in divided doses every 6 hours, and treatment continued until day 10. Since the therapy was initiated at

any time within 48 hours of the bleed, this resulted in 8-10 days of dosing, depending on when the medication was actually started.

[Slide]

Before I get into the actual results of the studies, I would like to discuss the neurograde in some detail. If approved, the proposed target population is subarachnoid hemorrhage patients with poor neurologic function at because following the initial hemorrhage. Poor neurologic function was defined in the studies using the neurograde.

[Slide]

The determination of the neurograde is based on a modification of the traditional Glasgow Coma Scale. I show the GCS here but I won't describe it since it is a familiar scale. Suffice it to say that the traditional GCS consists of the sum of the best scores obtained in all three components of eye opening, verbal and motor response, and it ranges from a low score of 3 to a high score of 15.

[Slide]

The modified GCS as defined in the study reports was determined at baseline by recording the patients' eye opening score, verbal score and the four individual limb motor responses. Unlike the traditional Glasgow Coma Scale which uses the best responses, the modified GCS was

calculated by using the worst motor response in the calculation of the total score. This modification was used because this may be a better predictor mortality and good outcome.

[Slide]

An imputation algorithm was used if there were missing components of the modified GCS. When this occurred the verbal score was the most likely one to be missing due to intubation. In this case, a verbal score of 1, the lowest possible score, was imputed.

[Slide]

The neurograde was assigned based on these cut-off scores from the modified GCS. Scores of 8 and below placed one in the high neurograde IV/V subgroup. It is this latter subgroup for which tirilazad is intended. Throughout my talk I use the term high neurogrades synonymously with neurogrades IV and V, and the term low neurogrades for I, II and III.

[Slide]

This slide compares the various neurogrades defined by the modified GCS and the Hunt and Hess Scale, which is the scale used in the nimodipine trials. Nimodipine is indicated in patients who have a Hunt and Hess of 1, 2 or 3. Although we can agree that increasing Hunt and Hess grades and increasing neurograde grades are both associated

with poor neurologic function, it is not obvious how well the individual neurogrades and Hunt and Hess grades correlate with each other.

[Slide]

There are three improvement effectiveness endpoints. Which one was the primary endpoint depended on the study. Mortality was assessed at 3 months, and 3 months was defined as any time between days 76 and 106. For studies 32 and 29 the sponsor analyzed mortality at day 76, and for studies 65 and 53 mortality at day 91 was used.

The Glasgow Outcome Scale, a measure of functional outcome which I will describe further, was assessed at 3 months. The presence of clinical vasospasm at any time during the treatment period was also recorded. The treatment period in this case was defined as the 14 days following the first dose of study medication.

[Slide]

This is the Glasgow Outcome Scale. It is a 5-point scale, as shown on the slide. Good recovery is a 1 and it ranges all the way to death which is a 5.

[Slide]

Other efficacy endpoints included the need for hypertension, hypovolemia and hemodilution therapy, neurologic worsening from vasospasm, and cerebral infarction during treatment.

[Slide]

There were some important differences among the studies. Studies 32 and 29 included both men and women. There were 4 dose groups in study 32, including vehicle placebo. Study 29 dropped the lowest tirilazad dose group. And, both studies 65 and 63 included only women and had 2 dose groups, vehicle and 15 mg.

[Slide]

Of the four studies, study 32 was the first study completed. It was conducted in Europe, Australia and New Zealand. It enrolled both men and women, and it treated 1,015 patients. There were four randomized treatment groups, as shown here, placebo, 0.6, 2 and 6. All patients received nimodipine. Throughout my talk this morning I use PBO on my slides as an abbreviation for vehicle placebo. This study was negative on its primary endpoint which was vasospasm, although a numerical trend in favor of tirilazad was seen.

[Slide]

I would like to say here that the primary endpoint was tested at the 0.05 level of significance with adjustment for multiple dose group comparisons. In general, acceptance of any other positive secondary or retrospective analysis inflates the type-1 error of the experiment above this 0.05 level. As we discuss secondary endpoint results from this and other studies, an improvement question to consider is

how much inflation of the type-1 error we are willing to accept.

[Slide]

The study also looked at mortality as a secondary endpoint. Mortality in the highest dose group was decreased compared to placebo. The comparison between 6 mg and placebo had a nominal p value, shown here, of 0.01.

[Slide]

This table subdivides the mortality data by sex. On further analysis one can see that the overall effect on mortality came entirely from the effect in men, as shown here, in this row, 2 percent for drug, again the highest does, 25 percent for placebo. As you can see, there was no between group difference in mortality in women. One possible explanation at the time was the higher clearance of the drug by premenopausal females which led to the use of a higher dose in studies 65 and 63. Analyses of other secondary efficacy measures were all negative.

[Slide]

Study 29 was conducted in the United States and Canada. It also included men and women and it treated 897 patients. This had 3 randomized treatment groups, placebo, 2 mg and 6 mg, and all patients received nimodipine.

[Slide]

The primary endpoint for study 29 changed while

the study was in progress. Initially it was vasospasm and GOS. Then it was vasospasm alone. But when the results of study 32 became known the sponsor and the Division held a meeting in late January, 1994, while the study was still in progress, during which time it was mutually agreed that mortality would be analyzed.

[Slide]

This table shows the mortality results for that study. I only show the data comparing the highest dose, 6 mg, and placebo. The first row shows the mortality data for the entire study population. The next two rows show the data for each sex. There were no statistically significant between group differences in mortality, and the study failed to replicate the positive mortality effect that was seen in men in study 32.

[Slide]

The sponsor then performed a retrospective analysis of the data for high neurograde patients. Now, we have no evidence to suggest at the time that this was a prespecified subgroup analysis, but in this severely ill group the difference in mortality was nominally significant at day 76 only in men treated with the highest dose. As you can see, 1/20, 5 percent, on drug; 4/12, 33 percent, on placebo.

I would like to point out the very small size of

this subgroup, 32 patients out of a total of 897 patients who were treated. But this was the first indication that tirilazad may have a treatment effect on high neurograde patients only.

The Division objected to the retrospective nature of this analysis of a very small subgroup, and also objected to the use of the day 76 data, the beginning of the 3-month window, since this analysis may have ignored any additional follow-up information that may have been available. We performed an analysis at day 91, and one can see that the inclusion of a single additional death in the high dose, from 1/20 to 2/20 -- this death occurred on day 84, results in loss of nominal significance, and the p value is even higher with adjustments for 2 doses, 2 genders and the 2 neurograde subgroups. The analyses of other efficacy measures, including the 3-month GOS, were all negative in this study.

[Slide]

In the June, 1995 non-approvable letter the agency considered the positive mortality effect in men seen in study 32 a statistically robust finding even though that study was negative on its primary measure. We also concluded that study 29 did not replicate this finding. Therefore, there was insufficient evidence for approval. The letter also stated that evidence of efficacy in women, along with

an increase in favorable outcome, would provide corroborative evidence needed to establish efficacy.

[Slide]

In July, 1998 the sponsor submitted a response to the non-approvable letter, and this submission contained the results of the 2 new large studies in women only, number 65 and 63, using a higher dose, 15 mg/kg/day.

[Slide]

Study 65 was conducted in Europe, Australia and New Zealand. It enrolled and treated 817 women. They were randomized to placebo or 15 mg, and all patients received nimodipine.

[Slide]

The primary endpoint in this study was mortality at day 91, and the results of this analysis are shown in the first row of this table. There was no statistically significant between group difference seen.

The sponsor also performed a retrospective analysis of mortality in the low and high neurogrades and there were no nominally significant between group differences seen in either of these subgroups.

[Slide]

Here is a graphical representation of the same data. In this chart the blue cylinders represent the overall study population, showing essentially no between group

difference in mortality. There was a numerical trend in the high neurograde group in favor of tirilazad, which is reminiscent of the similar trend that was seen in the high neurograde in men in study 29. There was a slight numerical trend in favor of placebo in the low neurogrades.

[Slide]

There were two analyses of secondary endpoints that were nominally significant. Clinical vasospasm and death from clinical vasospasm were decreased in the tirilazad group. However, this finding did not translate into any other measurable benefit such as decreased overall mortality or decreased incidence of cerebral infarction, and the analyses of other secondary endpoints were negative, including functional outcome.

[Slide]

My conclusions of that are this study failed to demonstrate a between group difference in mortality in women. There was no improvement in 3-month functional outcome. And, there was a decreased incidence of clinical vasospasm but this did not translate into any demonstrable improvement in mortality, functional outcome or incidence of cerebral infarction. This was essentially a negative study in women, and it failed to provide the corroborative evidence needed for approval.

[Slide]

I move now to the results of study 63 which was conducted in the U.S., Canada and Mexico. This treated 823 women and, like study 65, there were 2 treatment groups and, as in the previous 3 studies, all patients received nimodipine.

[Slide]

The original protocol specified that primary efficacy analysis was mortality at day 91 in the overall population. However, because of the favorable numerical trends in mortality seen in previous studies in neurogrades IV and V, the sponsor filed an amendment on December 16, 1996 which changed the primary efficacy analysis to mortality in neurogrades IV and V. This change occurred just before study completion, after enrollment of the last patient and before breaking the blind.

[Slide]

This slide shows the effect of the change in the primary analysis population on the sample size. Of the 823 patients treated, there were 154 patients in the high neurograde subgroup, as shown in this purple wedge. This represented 19 percent of the overall study population. Of these, 69 received tirilazad and 85 received placebo.

[Slide]

The primary endpoint in the study was mortality at day 91 in the high neurograde patients. This analysis was

positive, as shown in the last row of this table, showing a statistically significant between group difference in mortality in favor of tirilazad, 24.6 percent in the tirilazad group and 43.4 percent in placebo.

The sponsor also analyzed mortality in the entire study population, in the first row, and in the low neurogrades. These were both negative. Notice the numerical trend in favor of placebo in the low neurogrades just as was seen in study 65.

[Slide]

Here is a graphical representation of the same data. The yellow cylinders here represent the primary analysis population, high neurograde patients, and the tirilazad-associated reduction in mortality is evident from the graph. Mortality in the overall population was numerically lower in tirilazad but it was not statistically significant. As you saw in the previous table, the mortality was numerically higher in the tirilazad-treated low neurograde patients but also was not significant.

[Slide]

I would like to discuss the results of the GOS in this study. There is no universally agreed upon method to analyze the GOS; many approaches have been used. The sponsor chose three analyses which compared binary groups. The good recovery analysis compared the proportion of patients that

achieved a good recovery or a GOS of 1. The favorable outcome analysis compared the proportion of patients that achieved a GOS of 1 and 2 between the 2 groups. And, the vegetative/death analysis compared the proportion of patients that had a GOS of 4 and 5.

[Slide]

This table shows the results of the GOS at 23 months in high neurograde patients. As you can see, the odds ratios were all less than 1 which favored tirilazad. The composite analysis, here in the first row, analyzed the results of all 3 individual binary analyses simultaneously. This composite analysis and the analysis of vegetative/death were nominally significant. Since there were few in the vegetative category, most of them had a GOS of 5, the favorable results in GOS were generally a reflection of the decreased mortality. Therefore, the GOS result really offers no new independent findings and is merely a reflection of the positive mortality results seen previously.

[Slide]

Here is the GOS analysis for the overall population and for the low neurograde patients. The results were negative for the overall population, with odds ratios all close to 1. For the low neurograde patients there was a numerical advantage to placebo in all analyses and it achieved nominal significance in the favorable outcome

analysis. Put another way, the use of tirilazad in low neurograde women in this study was associated with a nominally significant lower proportion of favorable outcomes, GOS 1 and 2, compared to placebo.

[Slide]

There were no nominally significant between group differences seen in any of these other secondary endpoints.

[Slide]

In conclusion, study 63 demonstrated that tirilazad therapy was associated with a statistically significant decrease in mortality in high neurograde women. The question remains whether this is due to a drug effect and I will explore this issue further in my discussion.

There was also an improvement in functional outcome with tirilazad, however, this was largely a reflection of the effect seen on mortality. And, functional outcome was worse in low neurograde women treated with tirilazad, and mortality was at least numerically increased in this subgroup as well.

[Slide]

I would like to proceed now with my discussion of the data. My discussion focuses on four questions: Number one, is there substantial evidence of efficacy? Number two, can clinicians identify the target population easily and accurately? Number three, is there a risk of treatment to

low neurograde patients? And, lastly, what is the effect of concomitant nimodipine in the high neurograde patients?

[Slide]

I would like to review the progression of events during development that has led us to the first and most improvement question, is there evidence of efficacy in high neurograde patients?

In the non-approvable letter the agency acknowledged that there existed evidence of efficacy in men. This was due to the positive effects on mortality seen in all men in study 32, even though that study was negative on its primary endpoint. I would point out that efficacy in neurogrades IV and V at that time was not yet an issue, and that subgroup analysis was not done. Out of interest, I will present these data shortly.

Study 29 was then completed and it was negative on the mortality analysis as well as the mortality analysis in men. However, a subgroup analysis showed a positive numerical effect on mort in high neurograde men.

[Slide]

At that point, the agency issued the non-approvable letter and requested evidence of efficacy in women. Studies 65 and 63 were done using a higher dose. Study 65 was negative on mortality but showed a positive numerical mortality effect in high neurograde women.

Finally, study 63, which was originally designed to look at mortality in all women, was then amended to look at mortality in nigh neurograde women as its primary analysis and this showed a positive mortality effect only in this subgroup.

I should point out that neither study provides the evidence requested in the non-approvable letter, that of efficacy in all women. So, is there instead evidence for efficacy in high neurogrades? Since the efficacy data for each gender comes from different studies using different doses, I will continue to look at each gender separately.

[Slide]

This slide summarizes the evidence for efficacy in nigh neurograde men. The subgroup analysis shown here for study 32 is new in the sense that it was not prespecified in the protocol and was not submitted as part of the original NDA.

It is presented here as a retrospective analysis since this is the focus on the IV/V population. I want to point out that the subgroups are very small. In study 32 we are talking about 34 patients out of a total of 1,015 that were treated. There were no deaths in the tirilazad group, and the p value is nominally significant. However, I remind everyone that it carries no inferential value because of the retrospective nature of the analysis, and we performed a

minimal adjustment by taking into account 3 dose groups, 2 endpoints, vasospasm and mortality, 2 genders, 2 neurograde subgroups. As you can see, the p value loses nominal significance with such an adjustment.

On the right are the results of the same analysis for study 29 which you have already seen. Again, we see very small numbers. In this study we are talking about 32 patients out of almost 900 that were treated. Again, we see a numerical trend in favor of tirilazad but the p value does not reach nominal significance, and with an adjustment it is even larger.

I want to emphasize again the very small numbers that we are dealing with here. An important question before us is can we draw a conclusion about the efficacy of the drug in high neurograde men using retrospective analyses of such small subgroups?

[Slide]

This table shows that the positive mortality effect seen in high neurograde men in study 32 was really not limited to that subgroup. It was also present in the much larger group of low neurograde men, 2.4 percent versus 18 percent. The point of this slide is to show that the positive mortality effects seen in all men in study 32 are not coming just from the high neurograde subgroup, and it is difficult to argue that this study somehow replicates the

results seen in high neurograde women in study 63 since the positive mortality effect in all men in study 32 is mostly coming from the larger subgroup of men with neurogrades I, II and III.

[Slide]

Here I show the same efficacy results for high neurograde women. Study 65 did show a numerical trend in favor of tirilazad but the p value failed to reach nominal significance. As you now know, study 63 showed a statistically significant between group difference in mortality in favor of tirilazad, with a p value of 0.016.

[Slide]

In order to better understand the mortality results just described for study 63, we asked the question were there baseline imbalances in this subgroup? I would like to repeat what Dr. Katz already mentioned, that we don't ordinarily do these types of analyses because of their retrospective nature but we felt compelled to do so in this case due to the unusual circumstance created by the selection of a relatively small subgroup as the primary analysis so late in the study.

We were particularly interested in this question because the original randomization of the study population was not stratified by neurograde subgroups. Our statisticians point out that this fact makes it more likely

for because imbalances and important prognostic factors to occur in subgroup analyses of non-random subgroups. In addition, this subgroup represented approximately 20 percent of the overall population, and the severe reduction in sample size also increased the chance of imbalances at baseline to occur.

[Slide]

We used a controlled population in study 63 to identify important baseline prognostic factors. We selected 4 factors of interest that were adversely associated with mortality in placebo patients. These are age over 65, the presence of intraventricular blood on CT, a thick clot on CT, and the presence of a poor bilateral motor response, abbreviated as PMR2. We selected this last factor because we wanted to find the factor that identified very sick patients within the already quite ill IV/V subgroup because an imbalance at baseline of these very sick patients would be of clinical interest. A patient was said to have a poor bilateral motor response if they had a motor response manifested by the decerebrate rigidity or worse, decerebrate posturing, on both sides of the body. This translated to a motor component score of the GCS of 2 or 1 bilaterally.

I should say that this is, by all means, not a complete list. There are undoubtedly other important risk factors that were recorded and others that we don't even

know about and could not possibly have been measured. However, my point by presenting this analysis is this, once you select a subgroup for analysis that was not properly randomized at study onset, and substantially reduce the sample size, then the risk for baseline imbalances of both known and unknown prognostic factors increase, often making interpretation of such subgroup analyses difficult even if they were prospectively defined prior to study completion.

[Slide]

This table shows that in the entire control population in the study of 413 patients these 4 risk factors were, in fact, associated with increased mortality. Of the 4, poor bilateral motor response had the highest risk ratio.

DR. GILMAN: Dr. Oliva, can we interrupt for a question?

DR. OLIVA: Yes.

DR. GROTTA: Just a quick question, I know of the first three as being clear-cut, well-known predictors of outcome. How did you arrive at the PMR2? Maybe I just don't know, but is that a known in the subarachnoid hemorrhage literature, and been validated as a poor predictor?

DR. OLIVA: No, this was something that we retrospectively constructed with discussion, just selecting some clinical sign that we thought clinically at least made some sense, that if you had the decerebrate posturing on

subgroup was 154 so now we are talking about an even smaller subgroup of 63 patients.

In fact, this slide shows the mortality in the subgroup of IV and V patients in patients with poor bilateral motor responses and those without. It shows that the apparent decreased mortality was coming entirely from the subgroup of patients with poor bilateral motor responses at baseline, 5/23 versus 23/40. I admit that these are small numbers but, in fact, they are actually larger than the numbers we saw for high neurograde men. Those in the IV/V subgroup without PMR2 showed essentially no reduction in mortality with treatment.

So, does this mean that tirilazad only works in a subgroup of IV/V patients, those with poor bilateral motor responses? Well, this certainly makes very little clinical sense. It is difficult enough to imagine why tirilazad would only work in high neurograde patients, but why is the signal only coming from an even smaller group of very ill patients that represents only 40 percent of the IV/V subgroup? I certainly don't have an answer to that question but just as the positive mortality results seen in study 32 came entirely from men and this raised doubts in our mind about the drug's effect in the entire study population, this analysis, at least to me, raises similar doubts about the drug's effect in the entire target population of high

neurograde patients.

[Slide]

To summarize the efficacy data, this slide shows the mortality data from the 4 efficacy studies in the high neurograde subgroups. Mortality was lower numerically in tirilazad for all 4 studies. The analysis of mortality was retrospective in 3/4 studies. The evidence of efficacy in men comes from 2 very small subgroups in studies 32 and 29. The nominal p value for 32, though significant, loses significance when adjusted.

In study 63, in the only prospectively defined analysis, there was a statistically significant reduction in mortality in favor of tirilazad. However, the interpretation of the results is limited by the fact that there were baseline imbalances of important prognostic factors. The placebo group was older and sicker than the drug group using PMR2 as a marker. Furthermore, the positive mortality effect seen in the study seems to be coming entirely from a yet even smaller subgroup of the very sick patients.

In summary, the positive mortality finding in study 63 has yet to be replicated, in my opinion, in a prospective, randomized, controlled trial in the intended treatment population.

[Slide]

Now, some of you may certainly disagree with my

interpretation of the efficacy data as I have described, but one thing that we cannot disagree about is the source of the data. There were a total of 3,552 treated patients in the 4 efficacy trials. The best results in men, whether or not you agree that the data support a drug effect in men, come from study 32. In that study there were 176 men exposed to high dose or placebo, and of these, only 34 were in the high neurograde subgroup.

For women the best results come from study 63 which treated 823 patients. Of these, 154 were in the high neurograde subgroup and, as I have shown, the mortality signal is coming from an even smaller subgroup of IV/V patients with bilateral poor motor responses.

So, out of 3,552 we are left with efficacy data from less than 100. I pose the question can we conclude anything about the efficacy of the drug based on such small numbers?

[Slide]

I would like to move on now to discuss the target population. Let's assume that tirilazad is effective in neurograde IV and V. How easy will it be in clinical practice to identify those patients who should and should not receive the drug?

The obvious answer is to apply the same neurologic grading scale that was used in the clinical trials to

identify the IV/V patients. I would like to point out that to our best knowledge the neurograde is not a standard scale that is widely known or used in the medical community. It is an innovative scale that was implemented in these studies because it was felt to be a better predictor of good outcome and mortality. Since the original intention of the drug development program was to demonstrate efficacy of tirilazad in all neurogrades, I don't believe it was ever the original intention that such a scale would need to be applied systematically after approval.

[Slide]

As I mentioned earlier, it is a scale which requires the calculation of a modified Glasgow Coma score based on the worst motor component using a rather complex imputation algorithm for missing components, those where the modified GCS of 8 or less fall into the high neurograde category. Patients who were intubated had missing verbal scores and an accurate neurograde determination was not possible. In that instance, an arbitrary verb score of 1 was imputed. Since many seriously ill patients are intubated, the neurograde scale as designed really does not allow an accurate neurograde determination in these patients.

[Slide]

Across all 4 studies 13 percent of patients had missing verbal scores at baseline, presumably most, if not

all, due to intubation. But of the patients in the IV/V subgroup almost half, 44 percent, had missing verbal scores.

[Slide]

This suggests that the neurograde scale as defined in the development program is not the best scale to apply to seriously ill patients because almost half of these classified in the IV/V subgroup had missing verbal scores. This undoubtedly led to misclassification of at least some low grade patients into the high grade.

Another point to consider, as Dr. Katz pointed out, is that nimodipine is approved for use only in Hunt and Hess grades I through III. Strictly speaking, a clinician would have to apply two scales, the Hunt and Hess to determine the need for nimodipine and the neurograde to determine the need for tirilazad. It is my personal opinion that in a busy emergency room or intensive care unit setting it is doubtful that the neurograde scale would be applied accurately and consistently, leading to widespread off-label use of tirilazad in low neurograde patients which, I would also like to mention, outnumbered high neurograde patients in these clinical studies by about 4:1.

[Slide]

This leads us to the third question for discussion, what is the risk of treatment to low neurograde patients? Ordinarily the misclassification of a patient

resulting in an inappropriate use of a medication is not a serious concern if the medication, at the very least, has no adverse effect on the individual. But in the case of tirilazad we do have some evidence from the efficacy data that low neurograde patients treated with tirilazad have a worse outcome.

[Slide]

Here I show again the 3-month GOS results in study 63 of low neurograde patients. It shows that those treated with tirilazad did worse, with odds ratios greater than 1 for each analysis. Mortality was also numerically, but not statistically, higher in this group as well.

This is the very same study, as you recall, which also revealed a positive mortality effect in the high neurogrades. Now, this trend is not seen in either 65 or 32.

[Slide]

But let's look at study 29. This is the only other study which showed a numerically lower mortality effect with tirilazad therapy only in the high neurograde men but not in the low grades. When one looks at the subgroup of low neurograde men treated with the high dose, one sees a similar trend.

[Slide]

This figure shows the distribution of GOS scores in low neurograde men in that study, study 29. The blue here

represents patients that have a GOS of 1 or good recovery. One can see that numerically a lower percentage of tirilazad-treated men had a good recovery compared to placebo. This was largely due to a shift from a good recovery into the next category, moderate disability, shown in the green. Notice that the vegetative/death group was also larger in this treatment group, but only slightly so. I only point this out to show that the worsening in one group was not accompanied by improvement in another.

[Slide]

This chart compares the proportion of men achieving a good recovery at 3 months, and it shows a lower percentage of good recovery in tirilazad-treated men. The nominal p value here was 0.05.

Now, just as one can argue that the post hoc subgroup analysis of mortality in the high neurograde groups is inappropriate, the same arguments can be made here with equal validity. Assuming the interpretation that tirilazad has an adverse effect on functional outcome in low neurograde patients is not true, then we must explain these findings in some other way. One other possible explanation that the results of study 63 and 29 resemble what one might expect from studies that are overall negative, where analysis of one subgroup goes in one direction due to chance and the analysis of the complementary subgroup goes in the

opposite direction. The only difference, of course, is that the analysis of the high neurograde subgroup in study 63 was the prespecified primary analysis.

[Slide]

I show this relatively busy chart to illustrate that the worsening in functional outcome seen in the low neurogrades in both studies 63 and 29 is at least numerically supported by the mortality data. This chart shows the relative risk of dying on tirilazad therapy by study, and subdivided by neurograde. The blue here represents I, II and III and the maroon is IV and V.

Studies 32 and 29 used the data for high dose men only. A risk ratio less than 1 -- so anything below this line favors tirilazad. The blue bars, again, are the low neurogrades. I drew in the black vertical lines to represent the 95 percent confidence intervals which, one can see, are usually very wide and include 1 in most cases. So, all we can really say is that these are numerical trends. The absolute numbers for each subgroup are shown below out of interest. Remember that there were no deaths in the high neurograde men from study 32, which is why the relative risk there is zero.

Anyway, one can see that the relative risk for high neurogrades is numerically less than 1 for each study, as I described earlier, however, the relative risks in 65

and 63 for the low neurogrades are greater than 1.

[Slide]

Finally, I would like to discuss the effect of nimodipine in this population of severely ill patients and what effect this may have on the interpretation of the data. As you know, all patients in all neurogrades received concomitant nimodipine and, according to the nimodipine product labeling, it is approved for Hunt and Hess grades I through III only. So, this raises the question in our minds what is the effect of nimodipine on mortality in the high neurograde patients.

We don't know the answer to that, but there is some evidence in the literature to suggest that nimodipine may increase mortality in patients with poor neurologic grades. I am referring to a paper that was published in The Journal of Neurosurgery in 1998, describing the results of a Canadian study. This study is also described in the product labeling for nimodipine.

[Slide]

This was a randomized, double-blind, placebo-controlled multicenter trial which enrolled 188 patients with Hunt and Hess grades III through IV at baseline. The dose used was 90 mg every 4 hours which is higher than the recommended dose of 60 and may have been too high. The primary outcome was a 3-month GOS and it showed

that a higher proportion of patients on nimodipine achieved a good recovery. They also looked at mortality.

[Slide]

Here are the mortality results from that study. Mortality at 3 months in the nimodipine group was 54 percent compared to placebo at 30 percent. Now, the study did not report a p value but our chi square and shows a nominal p value of 0.044.

To the extent that Hunt and Hess grades III through V patients are similar to neurogrades IV and V, there exists at least a possibility that the use of off-label nimodipine in these patients may increase their mortality.

We are also aware of the results from a tirilazad plus nimodipine interaction study in animals which was submitted with the original NDA. This study suggested the presence of an adverse interaction between the two drugs. In that animal study the beneficial neuroprotective effects of either drug used alone on the hippocampus and lateral cortex paradoxically decreased when the two drugs were used in combination.

[Slide]

So, what can we say about the combination use of tirilazad plus off-label nimodipine in high neurograde patients? Well, not much at this point, except that there

exists at least the possibility that tirilazad reverses the adverse effect on mortality of nimodipine in high neurograde patients but that the overall mortality may still be higher compared to a true placebo without nimodipine. Since a trial placebo arm was missing from all the studies, this possibility cannot be excluded.

[Slide]

In summary -- and this is my last slide -- I close my talk by again showing the questions I raised during my discussion. For question number one, is there substantial evidence of efficacy? I have presented the efficacy data for the four large multicenter trials, with an emphasis on effects on mortality in the high neurograde patients. In all four studies there was a numerical mortality advantage of tirilazad over vehicle. In only one trial, study 63, was the study positive on its prespecified primary designated endpoint. However, in this study we found baseline imbalances of important prognostic factors in that subgroup, and the positive mortality signal appeared to come from an even smaller subgroup of more severely ill patients.

The other three studies rely on retrospective analyses of the high neurogrades. Since studies 29 and 65 were both negative studies, the positive numeric trend seen in men from studies 32 and 29 come from very small subgroups. Study 32, though nominally positive in high

neurograde men, was negative in this subgroup when we apply minimal adjustments to the p value.

For question number two, can clinicians identify the target population easily and accurately, I discussed the neurograde scale and the difficulties associated with identification of the high neurograde population using this scale, as well as the high incidence of missing verbal scores in the IV/V subgroup. Since the low neurograde group was roughly four times larger, I suggest that misclassification of patients in clinical practice would likely lead to widespread off-label use of the product.

For question number three, is there a risk to low neurograde patients, I described the possible risks to off-label use in the much larger group of low neurograde patients by examining the unfavorable 3-month functional outcome results from studies 63 and 29, and showing the numerically higher risk ratios in 65 and 63.

For the last question, what is the effect of concomitant nimodipine in high neurograde patients, I discussed the possibility that nimodipine may increase mortality in these patients and that the combination of tirilazad plus nimodipine may not produce any overall mortality benefit compared to a placebo arm without either medication.

This concludes my presentation. Thank you very

much for your attention.

DR. GILMAN: Thank you. That was a model of clarity, as was your presentation in the book that we read. Any questions? Dr. Brooke?

DR. BROOKE: I wonder if I can make a comment about mortality, which is obviously a very useful marker and became popular with the large cardiovascular studies. I am going to turn to another illness, actually, in amyotrophic lateral sclerosis there is a drug available which will prolong life but doesn't improve function. So, many of the people faced with this choice refuse the drug because they say, "why would I want to live for a longer period of time like this?" And, I wonder about the quality of life. I know there was no quality of life measure in these studies but I think we should keep in mind -- we have been arguing about whether mortality is or is not affected but prolonging the life of someone who is extremely disabled and dependent upon the hospital to exist, we should just bear that in mind when we are looking at these deaths. So, I don't think that prolonging mortality by itself is per se a good thing. It may be a good thing but it isn't per se a good thing. We should just bear that in mind.

DR. GILMAN: Dr. Van Belle?

DR. VAN BELLE: Yes, I would second the comment -- this was an excellent presentation. With statistics there is

usually a distinction between exploratory analysis and confirmatory analysis, and I think a lot of the analyses that were done by the sponsor would come under the category of exploratory in the sense that they were post hoc and not prespecified.

In fairness to the sponsor, if you go back now and consider the reanalysis that you did of study 32 with the high neurograde patients, since this was completely unanticipated at that time, would you consider that result to be exploratory or confirmatory?

DR. OLIVA: Well, it was completely retrospective, not specified in the study so I think it would meet your definition of exploratory, but we felt we had to look since that was the target population for which the drug would be indicated.

DR. VAN BELLE: I would not agree with that. I would say that it would be confirmatory in the sense that you would consider the initial exploratory analysis in the later study to be exploratory and, since this was not anticipated, I would classify it as confirmatory.

DR. OLIVA: Well, I guess I would have to agree with you on that point since it was done temporally after we had seen the results for the later studies.

DR. GILMAN: Why is that important for us to determine?

DR. VAN BELLE: Well, I think there are enough problems with the study that we should not find problems that aren't there. In other words, I think that going back to the original study and doing an analysis that was not anticipated or planned at that time, and to find the same result at that time, I think has a little bit more inferential strength than just doing repeated analyses without any prespecified hypotheses.

DR. GILMAN: But that then really augments the finding that it is a small group that accounted for this nominal significance.

DR. VAN BELLE: Right, and I think that in terms of the overall pattern there are still enough issues to be discussed, as we will do when meet with the sponsor but, in fairness, in terms of a category of exploratory-confirmatory, I would lean more towards a confirmatory result.

DR. OLIVA: I think we have a comment from our statistician, Dr. Cui.

DR. CUI: I am the statistical reviewer of Freedox. To me, study 29 for males is sort of exploratory because in this study the prespecified primary endpoint is vasospasm and the finding for the males basically is a post hoc analysis. When you go to the small high neurograde patient group, to me, it is exploratory, post hoc in nature.

DR. VAN BELLE: I am not arguing that it is post hoc. I am thinking about the label exploratory-confirmatory.

DR. GILMAN: Just to make this distinction clear, if the two statisticians disagree, exploratory would carry less weight. Is that the idea?

DR. VAN BELLE: Right. You know, when the FDA talks about two independent studies that is at least in part what they have mind. The second study confirms the results from the first study.

DR. GILMAN: Dr. Temple?

DR. TEMPLE: Remember, although 32 didn't meet its primary endpoint of vasospasm, it met the secondary endpoint overall, without subdividing, for the whole group, that turned out to be carried entirely by men. So, the subanalyses of men and subanalyses of very sick men are all attempts to look within and overall sort of positive study.

But I think the main reason for doing that is because all the rest of the data don't show anything in any overall group of males or females, but show what effect there might be in only a subset. So, going back to that original study in men, 32, to look at the specific subgroup that might be positive in the others is an attempt to see if you are being fooled. For example, if all of the action was in the I through III group, well, that would make even less sense than it already does, and I think that was the purpose

of it. So, I mean, all of these things are exploratory but the real reason was to go back and look at where the data were in the women and see if you could find the same sort of thing in the men in what was, however, a basically positive study once you buy off on the change in endpoints from vasospasm.

DR. CUI: May I say something?

DR. GILMAN: Dr. Cui, yes, please.

DR. CUI: This is just regarding what is a confirmative trial. For a confirmative trial usually we require a well-conducted trial, randomized, prespecified endpoint. So, if you do the study and have a significant finding you can attribute that finding to the treatment difference for the drug. That is a confirmative trial. But, say, for study 32, yes, we see sort of a trend, that is right but this study doesn't satisfy all the conditions for a well done confirmative trial. That is my point.

DR. GILMAN: Well, that is clear. Dr. Drachman?

DR. DRACHMAN: You found that in study 73, I think, the very riskiest patients, those with bilateral motor defects, did the best. Now, this is sort of counter intuitive. My guess would have been that when you looked at the balance this group would have been somewhat better. In other words, if the drug failed to work, then you would assume that the non-randomized imbalance would have favored

the group that is better. What is your explanation for why the tirilazad group with the bilateral decerebrate clustering did better?

DR. OLIVA: That is a very excellent question and it is one that I have thought about considerably, and I am sorry to say I don't have an adequate explanation, other than, you know, we are subdividing the data into smaller and smaller groups and at some point I think we just start seeing noise; we just start seeing chance findings that have no easy interpretation. I don't know if you have any other thoughts on that.

DR. CUI: Actually, this was done by me at the very beginning. I first felt the motivation to do that. The first thing is the nature of the subgroup analysis for these high neurograde patients. I assume you understand that if you do that without stratified randomization you incur the chance of imbalance in the baseline for prognostic factors. That made me worried; I wanted to check something. If I want to check something I want to find some indicators to check something. I know there may be many things, even some factors not measured in this study.

Now, I have some questions regarding that neurograde classification. I think it is too rough for neurograde V patients or it has missing verbal responses to impute the score for the verbal response of 1. Okay? But if

in these patients other things are okay, then these patients are classified to a higher neurograde but may still have a much better situation. If one person is only paralyzed in one arm three limbs are okay, but this person is categorized as a neurograde V, but to compare this person with a person paralyzed in all four limbs, I think this is quite different. So I want to explore the nature of this kind of neurograde classification so I want to see what happened with the really sick patients.

I also think this somehow addresses the question related to the overall trend that we find for the studies. Basically, for the low neurogrades the drug tended to have a worse mortality outcome, but for high neurograde, sicker patients the outcome is inferior of the drug. Now I want to see if there is this trend if you identify even sicker patients. So I did that and we see the trend. So, at that time I thought a lot about how to explain that. I asked Dr. Oliva and he said there seems no biological explanation for that.

Then I started to worry about the effect of nimodipine and I talked to Dr. Oliva about nimodipine, and he happened to mention that nimodipine was only approved for neurograde I to III patients. So I don't know how to explain that, but does that attribute to the effect of nimodipine or attribute anything else? We don't know. This is the finding

but the point is that anywhere where you find that there is a strong interaction between the treatment and the so-called PMR2.

DR. KATZ: I have a question and a comment. First of all, did you look at the PMR2 subgroup across studies?

DR. CUI: Yes, I did the same thing for study 65. The nominal p value for the overall neurograde IV and V is about 0.4, something like that. But if you restrict it to the very ill patient group with PMR2 the p value is significantly dropped to 0.08. So, the trend is similar.

DR. KATZ: The comment was, Dr. Drachman, you said that these PMR2 patients did better. It is not exactly clear what better means. If there is a drug effect, it seemed to be coming from well. How well they actually did is something to look at.

DR. DRACHMAN: But these are the ones who survived better. These are just the ones I would not have expected to have better survival.

DR. TEMPLE: You may want to look at the data. I don't think they actually survived better; they had a bigger drug effect. That is different. Actually, I was looking at that. I think they got up to a survival that was almost as good as the people who weren't in that category on treatment, and they were considerably worse when they were untreated.

DR. GILMAN: That is correct.

DR. TEMPLE: So it is where the drug effect might have been, if there is one, it is not that they did better. They responded better arguably because they had more to gain.

DR. KATZ: But it still begs the question as to how you explain this from a biological point of view, if that is necessary.

DR. CUI: I have one more comment. Usually when you see this kind of treatment by prognostic factor interactions, it is very hard. It imposes a difficulty to interpreting the oral finding. Okay? In this case, if you approve the drug for high neurograde IV/V patients -- suppose this is true, then you actually treat with the drug about 40 percent of the patients. For 60 percent of the patients the drug has no effect. That is the problem. If for 60 percent of patients the drug effect is there but for 40 percent of patients with no effect I would feel better.

DR. GILMAN: Any other questions from the committee for Dr. Oliva or Dr. Cui? If not, it is 10:20 and I would suggest a 15-minute break. Let's resume at 10:35.

[Brief recess]

DR. GILMAN: Let's resume again. I want to apologize to Dr. Oliva for mispronouncing his name at the end of the last session. Dr. Katz wanted to make a comment

before we begin.

DR. KATZ: Thanks. Yes, before we get into the safety just a couple of comments based on other comments at the end of the efficacy discussion.

About Dr. Brooke's comment about looking for meaningful survival, if you will, if you actually look at the non-approvable letter we said that what we wanted to be shown in the next study that would, hopefully, replicate the finding in study 32 was an effect on mortality and favorable outcome. So, I think we were cognizant of the fact that simply decreasing mortality at the expense of, you know, increasing the number of vegetative patients would be meaningless. So, we thought about it and I think we can talk more about that.

The other point I want to make is a difficult one to make and it has come up several times, and I am going to try and clarify what we thought had been shown prior to this resubmission vis-a-vis the effect of the drug in men because there is some sense that the advisory committee thought or the agency thought that we had shown that the drug had been shown to be efficacy in men. To the extent that that is an important point, I think it is useful to sort of clear the air about that.

I do not believe that the agency had concluded that the drug was effective in men. Had we concluded that, I

believe we might have approved it for use in men. What we believe -- and there may have been people on the advisory committee who believed otherwise but what we believed was that there had been a showing of statistical significance on mortality in one study in men, which is not the same thing as saying that the drug had been shown to be effective in men. We ordinarily require replication in order to conclude that and, in fact, to conclude that the statistical significance that was seen was attributable to the treatment. Without replication you are hard-pressed to say that. So that is really what we had concluded, and we needed replication. We chose to allow the sponsor to provide that replication in a study in women. You had raised the question about whether or not we had looked at in men. That is a different question. How you replicate it we can discuss. We did what we did. But the point is that we had not concluded that it was effective in men. We concluded that there was a statistically significant difference in one trial in men and that it would need to be replicated before we were willing that it was treatment related or effective.

DR. TEMPLE: And I would go one step further. The reason that study was persuasive was that the effect was present overall, not just in men, although when you looked it turned out to be driven entirely by men. So, the reason for asking for the second study in women was to overcome the

nagging suspicion that this worked only in one gender, in one sex. But it was to get another piece of evidence that it actually worked. Maybe that doesn't matter.

DR. GILMAN: Well, the basis of my question was that you had one positive and one negative study as far as I was concerned, just retrospectively looking at the data. That is why I asked the question.

All right, let's turn to safety issues. Dr. Judith Racoosin, medical officer, will discuss these.

Safety Issues

DR. RACOOSIN: Good morning. I would like to give the same slide disclaimer as Dr. Oliva gave. My presentation will be slightly different from the copy of the slides that you have.

[Slide]

This morning I will be focusing my comments to safety issues pertaining to the approvability of tirilazad. In doing so, my discussion will be limited to mortality and selected serious adverse events.

[Slide]

I will begin by elaborating on the mortality differences across neurograde strata presented by Dr. Oliva. I will then discuss mortality findings in non-pivotal subarachnoid hemorrhage studies. Next I will review the mortality experience in the acute ischemic stroke and head

injury indications. It is important to review this data because development of tirilazad was halted for both of these indications due to tirilazad-associated mortality excesses observed in large Phase III trials.

[Slide]

Before proceeding, I would like to discuss the limitations of the safety review. First, as you well know, subarachnoid hemorrhage patients are complicated. In each patient any combination of many different intracerebral processes were going on simultaneously. These included the direct effect of the initial bleed, cerebral edema, vasospasm, cerebral infarction, rebleed and angiographic and/or intraoperative complications. Furthermore, many patients had cardiovascular or pulmonary complications.

However, the review of such complicated patients could have been made easier by the sponsor by providing more informative narratives. Unfortunately, the narratives provided by the sponsor consisted only of a summary of the case report form. It was difficult to understand how the events related in the narrative related to each other in time, and a few details explaining these events were included.

Let me elaborate on that last point. In the whole NDA little, if any, supporting data was provided for the interpretation of deaths, discontinuations and adverse

events. Let me give you some examples. Patients who had serious adverse events called cerebral edema or intracranial hypertension did not have CT scan data or pertinent clinical data provided to explain that finding. Patients with ARDS had no chest x-ray data or Swann-Ganz catheter or pressure data. Patients with pancreatitis or thrombocytopenia had no pertinent lab values included in the narratives. This lack of data limited our ability to evaluate which events might be drug related and which might be related to the patient's underlying condition. Please keep this in mind as I discuss these issues this morning.

[Slide]

This table summarizes the mortality rates by neurograde for study 63 and 65. As you know, the endpoint for the efficacy analysis was 3 months. In my safety review, however, I used a study day 20 cut-off for measuring the frequency of death and adverse events. I chose this shorter period of time because in most cases the more proximal an event is to the drug exposure the more likely it is to be related to it. Day 20 was specifically chosen because the drug has a long half-life of 61 to 120 hours. So, it was likely that the drug was still present in the patient's system 10 days following the end of the infusion.

In the table you can see the mortality risk for vehicle and tirilazad for the low neurograde patients and

high neurograde patients in study 63 and 65 at 20 days and 91 days. These 91-day values are the same values that Dr. Oliva presented. As you can see, although the percentages are understandably smaller at 20 days, the relative risks are very similar to those seen at 91 days.

As before, the low neurograde patients who were treated with tirilazad were at an elevated risk of mortality and the tirilazad-treated high neurograde patients were at a decreased risk of mortality. Now, 1.3 may not seem like a huge relative risk but in mortality 1.3 correlates to a 30 percent increase in mortality. So, this is something that we would want to investigate. The p value on this is only about 0.3 to 0.4 in the 2 studies, what you would consider a weak signal but, again, in mortality we like to be certain of what we are looking at.

In trying to understand why this pattern of tirilazad-associated mortality occurred, I reviewed case report forms -- yes?

DR. GILMAN: Can we have a question?

DR. KAWAS: I just wanted clarification. You said 0.3 or 0.03?

DR. RACOOSIN: 0.3.

DR. KAWAS: 0.3 and 0.4.

DR. RACOOSIN: Yes.

DR. KAWAS: So, not significant.

DR. RACOOSIN: Correct, but let me just comment that in safety we don't generally use the same strict cut-offs that are used in efficacy to denote a statistically significant positive study with 0.05.

DR. KAWAS: I just wanted to make sure that we are talking about 0.4 now.

DR. RACOOSIN: Correct, 0.3 to 0.4.

I reviewed the case report forms and death narratives and analyzed the cause-specific mortality by neurograde strata. In reviewing the death narratives it became obvious that assigning one single primary cause of death in these patients was very difficult and also seemed somewhat arbitrary because so many different processes were going on at the same time. I did not find one single specific cause of death that explained the pattern of mortality risk in the low neurograde group.

[Slide]

Since the review of cause of death by neurograde was unrevealing, I then looked at treatment emergent adverse events, serious adverse events by treatment group and neurograde strata in all body systems. Let me comment here that a serious adverse event was denoted by the investigator. The regulatory definition includes life-threatening events, events that require hospitalization -- of course, these patients were already hospitalized --

and events that prolong the hospitalization.

For ease of analysis of adverse events the sponsor coded each investigator verbatim term into a broader category, called a COSTART term. The COSTART term in theory encompasses all the verbatim terms assigned to it. So, when I talk about COSTART terms today, for example, here, edema brain, the investigator verbatims were cerebral edema, brain edema and so forth but the sponsor categorizes them into a broader category called edema brain. I am also going to be discussing intracranial hypertension, the COSTART term for that, and that included increased intracranial pressure, cerebral herniation and those sorts of specific verbatim terms.

When I reviewed the frequency of serious adverse events I found that the occurrence of the COSTART term edema fit the same risk pattern as that of mortality shown in the previous table. Again, I want to reiterate that all we had to go on was the investigator's verbatim term related to edema brain. I can't tell you how many patients had this based on radiologic findings or clinical findings.

This table shows the occurrence or the percentage of patients who had the serious adverse event of edema brain in the vehicle group and tirilazad group at low and high neurogrades in study 63 and 65. Again, you can see that low neurograde patients in both studies were at an elevated risk

for having the serious adverse events although it is more marked in study 63. High neurograde tirilazad-treated patients appeared to have a decreased risk of serious adverse events. A similar pattern was seen with the COSTART term intracranial hypertension, although again the relative risks were not as marked.

[Slide]

I am going to now discuss the mortality experience in the Phase II trials. Study 0007 was a Canadian study and study 19 was a Japanese study. Both of these were early Phase II dose escalation trials. They differed from the Phase III trials on two important issues: one, the patients were allowed to be enrolled up to 72 hours and, secondly, the sickest patients, the neurograde V, were not enrolled -- they were excluded from enrollment. Additionally, patients in study 19 did not receive nimodipine.

Both studies were performed in a tiered fashion. So, the lowest dose was done compared to a vehicle group and the safety results were reviewed. Then, once it was observed to be safe, the next dosing level was initiated, and each dosing level had a placebo group with it.

When the mortality was compared in study 19 overall and it is not shown well here on this slide, but the overall comparison of mortality to vehicle in the treated group versus the vehicle group had an odds ratio of 6.5 for

mortality. The p value on this was 0.07. As you can see, in the female group in both studies there was a dose response for mortality, in both 0007 and study 19, and you can see that the tirilazad 0.6, high dose females, both had substantial mortality in that group. When the comparison was made of the high dose females in both studies to vehicle the difference was statistically significant. The p was 0.04 in this study and the p was 0.02 in this study. A similar risk for high dose women was not observed in the larger studies, 32 and 29.

[Slide]

Three studies were performed to compare safety across dose levels of tirilazad. These three studies did not have a vehicle control. Study 62 enrolled men only at the dose of 6 mg/kg/day versus 10 mg/kg/day. Study 55 included men and women with dose levels 6 mg/kg/day, 10 mg/kg/day and 15 mg/kg/day and study 56 enrolled men and women at 10 mg/kg/day and 15 mg/kg/day.

In study 55 there was an excess mortality risk in the highest dose group. So, the lower two doses had about 10 percent mortality and the high dose had 15 percent, and the p value for the comparison of 25 percent to 11 percent was 0.2. There was little difference in mortality risk between treatment arms in study 56 and study 62.

DR. GILMAN: Question, Dr. Grotta?

DR. RACOOSIN: Yes?

DR. GROTTA: This brings up a question I was going to actually ask later, and that is, I mean, if we accept the fact that men and women metabolize the drug differently, the dose of 15 mg/kg/day obviously is different in men and women. So, I guess the question I have is do we have evidence in men that 15 mg/kg/day is too high? I mean, these are data from men and women.

DR. RACOOSIN: Right, and in this particular study when we broke out the mortality by gender, both men and women contributed the same amount. So, in that high group there were 3/12 men who died and then there were 5/20 women. So, they contributed equally to that.

DR. GROTTA: I guess when we hear the company's presentation maybe they will address the question of how they arrived at 6 mg as the highest dose in men and didn't go higher in the development phase of the drug.

DR. GILMAN: Just to follow-up on that, it is premenopausal women in whom metabolism is higher, but postmenopausal women presumably not.

DR. RACOOSIN: In postmenopausal women there is only about 10-15 percent difference from what I understand, and these numbers are from single-dose studies. The premenopausal women had the much higher clearance as compared to men, and postmenopausal women had somewhat of a

higher clearance but only about 10-15 percent. But as we have come to understand in multiple dosing studies, especially when there is the addition of phenytoin in many of these patients, the differences in the clearance between gender become much smaller.

[Slide]

I would like to discuss the mortality experience in the acute ischemic stroke indication now. The efficacy and safety of tirilazad in acute ischemic stroke was evaluated in several trials, culminating in the Phase III studies 88, conducted in Europe and Australia and 81, conducted in North America.

Studies 81 and 88 used a dose of 10 mg/kg/day for men and 12 mg/kg/day for women for 12 doses. This was compared to a vehicle arm. After 355 of a planned 910 had been enrolled in study 88 the safety data monitoring board recommended termination of the study due to an increased mortality in tirilazad-treated subjects. Study 81 was terminated at this time as well. Only 126 of a planned 890 had been enrolled.

[Slide]

This table shows the mortality risk by time since study entry. So, these are the mortality risks for vehicle and tirilazad patients for study 88 and study 81 at 3 days, 5 days, 10 days and 3 months. The 10-day and 3-month time

periods -- these were calculated by the sponsor. The FDA did an additional analysis at 3 and 5 days.

As you can see, the greatest relative risk for mortality for tirilazad-treated patients is at 3 days, and this is right around the end of the infusion period.

DR. DRACHMAN: What did they die of? What was the cause of death? Do we know that?

DR. RACOOSIN: I am going to address that in the next slide.

As you can see, the relative risk does decrease over the course of the study, yet it is still elevated at the end, and you can see the risk difference between the 2 groups. It is about 3 percent here, and this is maintained over the course of the study, and it is about 4 percent at the end.

After the sponsor did an analysis adjusting for age and baseline neurologic status there was still a statistically significantly increased mortality risk associated with tirilazad exposure. This excess tirilazad-associated mortality risk was not observed in study 81, however, again, only about 50-odd patients had been enrolled in each treatment arm at that time.

[Slide]

Now getting to your question, death narratives were reviewed to gain insight into the tirilazad-associated

mortality excess observed in study 88. There was an excess of deaths. Here you can see for study 88 vehicle and tirilazad and these were the cause of death categories. There appeared to be an excess of deaths in the tirilazad group in the extension of admission infarct and in the hemorrhagic conversion. This wasn't observed in study 81.

DR. GILMAN: Another question.

DR. RACOOSIN: Yes?

DR. GROTTA: I would just like to make a comment. I mean, I don't know whose decision it was to stop that study, but it seems to me that those mortality figures are not all that bothersome. In fact, the mortality figure in the vehicle group was pretty low in the 88 study, and that may be the reason why a difference was seen. Actually, if you look at the various mortality rates, it is not that the tirilazad groups are high compared to common experience but in 88 the vehicle rate was low and in 81 the vehicle rate was extremely high. With these small numbers I am actually kind of surprised -- of course, I wasn't involved and wasn't there, but I am kind of surprised by the conclusion that there is a significant mortality difference, enough to stop a study.

DR. RACOOSIN: As I said, this was the recommendation of their treatment monitoring committee.

[Slide]

Now I am going to discuss the mortality experience in the head injury indication. Study 17, a North American study, and study 36, done in Europe, Australia and New Zealand, were multicenter, randomized, vehicle-controlled safety and efficacy studies of tirilazad in patients with moderate or severe head injury. Moderate head injury was defined as a baseline Glasgow Coma Scale of 9-12 and severe was 3-8.

Both studies used a dosing regimen of 10 mg/kg/day for all patients for 5 days versus a vehicle group. Study 17 was suspended on advice fm the treatment monitoring committee just short of study completion due to increased mortality in tirilazad-treated subjects. Study 36 was completed as planned, with a total of 1,131 patients.

[Slide]

Here we have the mortality risks in vehicle and tirilazad groups in study 36 and study 17 at 14 days, 3 months, 6 months and 12 months. Starting at 14 days, there was an elevated relative risk for mortality that was maintained over the course of the study. The p value for this was 0.055 and thereafter the p value was less than 0.05. When risk of mortality in study 17 was calculated by baseline severity tirilazad-treated patients that fell into the moderate designation had a higher mortality risk as compared to those in the severe group.

[Slide]

The sponsor did an analysis that identified differences in baseline factors between the vehicle and tirilazad groups. The sponsor asserted that after adjustment for imbalances in baseline characteristics no difference in mortality was found between the two treatment groups. This conclusion was based on a change in p value from a significant level to a non-significant level, up to a p of 0.11.

The FDA's repetition of this analysis confirmed that adjustment for baseline characteristics that the sponsor used, which were Glasgow Coma Scale, CT scan findings, pupil reactivity, systolic blood pressure and age, did raise the p value above 0.05. However, the FDA analysis also showed that while the p value changed the relative risk only changed about 10 percent, from 1.37 to 1.23, suggesting that the imbalanced distribution of baseline characteristics across treatment groups did not fully explain the mortality excess in the tirilazad group. Furthermore, as I mentioned earlier, the p value of 0.11 for mortality was still considered by us to be a safety signal.

We reviewed the death narratives in case report forms to try and identify a specific cause of death that might explain the mortality excess in the tirilazad group.

[Slide]

The primary cause of death was assigned by the investigator. The mortality excess in the tirilazad group in the study appeared to be primarily explained by an excess of deaths in the category related to elevated intracranial volume and, within that category, herniation.

You can see that 12 percent of the patients in the study died due to reasons relating to elevated intracranial volume, and these are the specific causes of death that were included in that category, and only 9 percent of patients in the vehicle group died of these causes. Within that, you can see for herniation 5.8 percent of patients in the study died due to herniation, who were treated with tirilazad, compared to 3.2 percent of the vehicle patients.

The 2 groups had similar numbers of herniations on study days 1 and 2, with 14 in the tirilazad group and 12 in the vehicle group. The difference in mortality due to herniation occurred between study days 3 and 7, with 13 in the tirilazad group and 4 in the vehicle group.

[Slide]

In order to get a sense of how unexpected it would be to observe two studies that had statistically significant mortality excesses associated with the drug among the non-subarachnoid hemorrhage studies that were conducted, we calculated the binomial probability of such an event. We included controlled studies that had adequate size to

demonstrate a difference in mortality. Although we treated each study as if it had an equal opportunity to demonstrate such a difference in mortality, this was probably not the case since the studies varied in size. The p value of 0.03, calculated by the binomial probability formula, suggests that it would be quite unexpected to observe 2/6 studies with a statistically significant tirilazad-associated excess mortality. This p value doesn't have inferential value; it is just calculated to get a sense of how unexpected it would be to make this observation.

[Slide]

I would like to conclude by summarizing the safety signals I have discussed this morning. There is the unexpected finding of two studies in subarachnoid hemorrhage indications with statistically significant tirilazad-associated mortality excesses. In both of these studies the mortality excesses appear to be related, at least in part, to a neurologic event.

Next, there are the Phase II subarachnoid hemorrhage studies, and 19 had a statistically significant overall drug-associated mortality risk, and both 9 and 7 had a statistically significant mortality excess in the high dose women group compared to vehicle -- the high dose treated group compared to vehicle.

How did these risks in these studies fit in with

the findings in studies 63 and 65? Those studies were done at higher doses of 15 mg/kg/day. As I described earlier, the Phase II patients did not include the sickest patients, the highest neurograde. As a result, the Phase II trial population was similar to the low neurograde populations in the later studies because the sickest patients had been excluded. As a result, the mortality excesses observed in the Phase II trials seem consistent with the mortality excess observed in the low neurograde group. These mortality signals are unexplained and unexpected, and must be factored into the discussion of the approvability of tirilazad.

Thank you. I would like to acknowledge Dr. Gerry Boehm, Dr. Joel Frieman, Dr. James Knudsen and Dr. Michael Sevka for their assistance in the safety review.

DR. GILMAN: Thank you, Dr. Racoosin. Could you answer some questions from the committee? First, as you look across all studies, are there complications that seem to come out from each of these studies that seem in common? For example, as I looked at your data it appeared that edema was one factor that was emerging I think pretty much across all studies because of the increase in intracranial volume that you pointed out. Second is the question of hemorrhage or the development of hemorrhage.

DR. RACOOSIN: Right, I do agree to a certain extent. I have to give one caveat though, and that is we

were very limited in being able to interpret the events. As I said, we were really lacking in information that we could use to sort of flesh out sort of just the name of the serious adverse events. So, I would agree, and it was the observation of the increase in herniation in the head injury studies that really led me to look at similar adverse events in the subarachnoid hemorrhage studies and I did see, you know, some trends. Overall there may not have been but, then, in the low neurograde group it appeared to be a little bit stronger.

DR. GILMAN: Can I just follow-up and ask how many autopsy verification studies were there to go along with these clinical observations?

DR. RACOOSIN: Let me say that in each case report form there was an indication of whether an autopsy was performed. However, it was rare that the autopsy report was included along with the case report form, especially in studies 63 and 65.

DR. GILMAN: So, you weren't sure whether an autopsy was or was not done?

DR. RACOOSIN: No, I could tell if it had been done but the results were generally not present.

DR. GILMAN: So, was autopsy often done?

DR. RACOOSIN: I would have to say that it appeared to be about 20-30 percent of the patients but that

is my general impression. I would have to go back and look to give you a specific answer on that. Maybe the sponsor could elaborate on that.

DR. GILMAN: We could ask the sponsor. But then, of that roughly 20 percent, how many were reported in the data you saw? The results of how many autopsies were reported?

DR. RACOOSIN: Well, I would have to say that I could probably count on fingers the number of autopsy reports that were addended in the case report forms. Now, I have to say I didn't read every single case report form but I reviewed 50 percent of the case report forms of the deaths. So, it generally was not present. We have had some discussion with the sponsor about trying to obtain those but, you know, with the limited time -- that may be something we would like to get for follow-up.

DR. GROTTA: In patients with cerebral hemorrhage and other very severe brain injuries 3-month and shorter mortality sometimes can be misleading; 6-month mortality may be more accurate. I would be interested in any data that you have or the company has on 6-month mortality since we are focusing so much on the mortality issue here. It may take a while for patients to die.

DR. RACOOSIN: I am sorry, this is the subarachnoid hemorrhage patients --

DR. GROTTA: Well, including the subarachnoid hemorrhage data but also in your data because 6 months in severe head injury patients as well as in hemorrhage patients -- we have seen in our patient population a change in relative proportions in different treatment groups from 3 months to 6 months, and you may rescue patients and then see delayed death either in patients who you have intervened in or not. So, what I am saying is that 6-month data would give you a more accurate long-term result, and I would be interested to know if you have seen any 6-month data or whether the company has any to present to us.

DR. RACOOSIN: I did not focus on the long-term mortality.

DR. BURKHART: If I could just comment about that point about long-term mortality, we really have a couple of problems when you think of how to relate an event to an acute infusion period, for example. I mean, if you start going too far out then you are going to start adding deaths that can't possibly be related to the exposure. On the other hand, you are quite right that some events may actually begin during the infusion that are delayed. So, you have a dilemma as to how to capture the events. So, we usually focus on fairly close proximity to the infusion. In this case, you do see a relative difference. So, I am not so sure that I would be interested in seeing 6-month data if it

looked a lot different than early infusion data unless I was sure that there was a delay.

DR. GROTTA: Yes, I realize the point of contamination with delayed things, but if you are interceding, let's say, with a cerebral hemorrhage and you do surgery and you salvage the patient temporarily but then they languish in a chronic care facility and then ultimately still die of their event, over the short-term you may not see a change in mortality but it catches up. And, we have found that some of that catch-up occurs between 3 and 6 months. So, at least to the point that cerebral hemorrhage or cerebral hematoma patients reflect also subarachnoid hemorrhage or closed head injury patients, which I think they may, you might find something by looking at 6-month data that would be relevant.

DR. RACOOSIN: Let me just comment that those bleeding events that you are referring to, if they occurred during the infusion or in the first 20 days, I would have examined them in my review of either serious adverse events or adverse events. I realize that that is different than the mortality issue but overall for rebleed and those sorts of intracerebral bleeding events I didn't see differences between the treatment groups within that 20-day period.

DR. GILMAN: Dr. Brooke has a question.

DR. BROOKE: Yes, just with regard to the 6-month

question. If there were late complications from some early effect of tirilazad you would expect the percentage to change between the placebo group and tirilazad group mortalities, to change over the course of time. It would widen. And, there was no evidence of that from your numbers. The percentage difference in mortality between the vehicle and the treatment group was the same for the three points you looked at, which was perhaps a little reassuring.

DR. RACOOSIN: In the stroke study.

DR. BROOKE: That is right.

DR. GILMAN: Dr. Drachman?

DR. DRACHMAN: Were there any non-cerebral adverse events, blood, lungs, heart?

DR. RACOOSIN: We did observe differences between treatment groups for certain events. I am hesitant to describe them at length here mainly because of the lack of detail I have for those events.

Let me give you one example. In one of the large subarachnoid hemorrhage studies in women there was an excess of a COSTART term for acute respiratory distress syndrome. In the other study there wasn't an excess of that; there was an excess of something called respiratory disease. When I looked at the respiratory disease verbatim terms, most of them were respiratory failure. Now, how do I know that respiratory failure and ARDS aren't the same thing? I don't.

And, there was also lung edema which could also have played into it, and since I could not define the events I didn't want to go into great detail about the differences between studies because I really wasn't clear what I would be talking about. So, I would like to refrain from discussing any of that further.

DR. GILMAN: Other questions or comments? If not, thank you very much. We will now move along. We are going to hear next the Pharmacia and Upjohn presentations. The introduction will be by Dr. Mark Corrigan who is Vice President of Global Clinical Development, Pharmacia and Upjohn.

Pharmacia and Upjohn Presentation

Introduction: SAH Development Program

Past and Present Issues

DR. CORRIGAN: Thank you Mr. Chairman, members of the FDA and the audience.

[Slide]

I have the opportunity to discuss with you the sponsor's brief presentation on Freedox, and in an effort to make the most salient points and keep the presentations brief we intend to follow the following order. I will briefly go through the development program and some of the rationale, much of which has been presented and so I will move quickly through that, past and present issues. At that

point Dr. Marshall will come up and focus on the points raised by the FDA. At that point we will respond to any questions that the committee may have.

DR. GILMAN: Let me comment to the committee that you have the slides in this book.

DR. CORRIGAN: To discuss the indication, I think it is worth noting that we believe we have met the FDA and congressional standard of substantial evidence for an effect for tirilazad mesylate in patients with subarachnoid hemorrhage to improve survival and functional outcome in patients with poor neurological function following the initial hemorrhage. Treatment should be initiated within 48 hours of initial hemorrhage, preferably prior to surgery and perhaps patients may show some advantage for early administration.

[Slide]

The entire Freedox program, as you have seen, has been conducted in over 10,000 patients in a number of diseases, including subarachnoid hemorrhage, ischemic stroke, spinal cord and head injury. Additionally, there was a large ongoing program for renal site protection.

The subarachnoid hemorrhage, identified in yellow, is the largest ever conducted in this disease entity, with over 3,000 patients who received Freedox.

[Slide]

I will review the program briefly for the committee who have already seen it. The first study that you have heard about was study 32 in which vasospasm for all neurogrades was the primary endpoint in keeping with the scientific understanding of the disease at that time.

As has been pointed out, although no effect was found in the primary endpoint, a significant effect on mortality was found in the study, essentially driven by the males.

[Slide]

As a result of this finding but prior to analysis, and in consultation with the FDA, the endpoint was altered for study 29 to mortality. It might be noted that there was no change was made in the entrance criteria in terms of the modified Glasgow from worst motor score, which may have been a better predictor in terms of vasospastic consequences, to best motor score which may correlate better with mortality outcome. While this study did not confirm the mortality events found in study 32 for all patients, a post hoc analysis of the most severe neurogrades indicated an effect, as you have heard, for those patients. Because of the paucity of treatment options for this population, the advisory committee meeting was held, the results of which you have heard.

[Slide]

Since pharmacokinetic differences may have driven the lack of effect in women, in consultation with the FDA, two further studies at higher doses in women were undertaken with mortality as the primary endpoint.

[Slide]

Study 65, the first study conducted in Europe and Australia, did not demonstrate an effect on the primary outcome, mortality, for all neurogrades. However, a predetermined subset of the most severely ill patients for the primary endpoint demonstrated a signal. Given the two previous studies, and prior to the completion of the 63 study, the primary endpoint for the second ongoing study in women was changed.

[Slide]

As you have heard, the final results for this study showed statistically significance for the primary endpoint of the effect of drug for the most severely affected women in the subgroup.

The process of clinical science for the development of novel therapeutics in disease entities, particularly first in class drugs are being shown, inevitably leads to a further understanding both of the disease process and of the agents being studied. To conduct investigations with full scientific disclosure and not avail oneself of the evolving understanding of the condition and

the agents under examination is to miss potentially effective treatments at best, and irresponsible science at worst.

[Slide]

The preponderance of the evidence for Freedox for subarachnoid hemorrhage is represented in the above summary histogram for the four adequate and well-controlled studies and the integrated summary is presented here. In all cases Freedox reduced mortality. Although the different doses that we are recommending in the product label are represented, therefore, males in 32 and 29; females in 65 and 63.

DR. GILMAN: Can I ask you a question about that?

DR. CORRIGAN: Sure.

DR. GILMAN: In your integrated bar graph are you comparing studies that had different genders, different doses? How can you show an integrated bar graph across all these studies?

DR. CORRIGAN: Since the entrance criteria and description of the disease state was the same for those studies in terms of the disease process, you are absolutely right. Genders are different. The different doses that we had are explained by the fact that we felt that the effect of 50 mg/kg was the correct dose in females.

DR. GILMAN: So, you are comparing apples and oranges here with respect to treatment administered in the

groups.

DR. CORRIGAN: Well --

DR. KAWAS: Could I add to that question? How did you integrate it? Did you just put all the subjects together, or did you act like each contributed equal -- how was the integration done? Maybe that is really what we want to hear.

DR. CORRIGAN: Well, this isn't a formal meta-analysis. This is an effort to look at -- yes, we basically just combined the patients.

DR. KAWAS: So, that is just an additive integration by pooling all the subjects in all the studies.

DR. GILMAN: Members of your team are saying no, that is not right.

DR. RUPPEL: Yes, excuse me, Mark. Betty Ruppel, biostatistics and data management for the company. We did not just pool all the patients to create a nuance ratio and estimates of the mortality rates. Instead, we individually analyzed each study with Cochran, Mantel-Henzel statistics and then combined those. So, it is a true meta-analysis rather than just a combination.

DR. KAWAS: Each study contributing equally?

DR. RUPPEL: Exactly.

DR. TEMPLE: Equally or were they weighted?

DR. RUPPEL: Well, they were weighted by the size