value doesn't make the nominal statistically significant cut point to say, well, that happened because we didn't have enough power. That's one possibility.

Another possibility is that there is no effect and we have to consider both. A nonsignificant p-value in a low-power test doesn't necessarily prove anything to me. If I said that wrong, say it right, please, Tom, or Dr. Koch.

DR. KOCH: Yes. I appreciate your concern but you should also appreciate that the patients who were not randomized at all had a magnitude of effect that was greater than or equal to the magnitude of effect that you're seeing in the warfarin subgroup. For the most part, the differences in the magnitudes here are all relatively small so there's really not much you can do with this.

DR. BORER: Okay. I don't want to belabor that.

DR. YUSUF: I want to give you some biological data, Jeff. One thing, just to look at this group non-randomized, because obviously if you

look at this group, you see that is almost identical to the overall result.

The next slide. It's worth noting that treatment with warfarin was initiated at a median of just under three days. Okay? And as we all know, it takes three to four days to get a warfarin effect. Now, this is showing you that the entire difference of warfarin occurred at 72 hours before you were able to get the warfarin effect.

Warfarin was just given about 2.8 days or so was when it was started. You know it takes three doses at least to get an effect. Biologically the entire difference is occurring before the warfarin is having an effect so this is just random chunks.

DR. BORER: Well, perhaps but, again, we're not going to resolve that here. Let me ask another question about your slide on -- in our slides it's page 48, OASIS-2, CVD/MI/RA absolute and relative benefit over time, intention to treat analysis, and the preceding slide where you show the deltas.

These deltas in absolute terms are very small and that's understood. As a percentage change

they are perhaps more impressive. The absolute values are small and I don't know that one can say anything that is statistically meaningful about the variation.

It appears to me that the deltas do get smaller over time. Now, you are looking at a seven-day endpoint. But in terms of the potential clinical utility of a therapy, if you know you're going to be losing benefit over time, then the strategy would have to be evaluated in the context of some additional management strategy to maintain benefit, I would think, or you might say that.

You dismissed the differences in these deltas as being basically insignificant or unimportant. I shouldn't use insignificant but unimportant. Yet, I see what looks like a trend to loss of benefit. I wonder what you think about that?

DR. YUSUF: I think conceptually on the fact that any short-term treatment must be followed by long-term treatment. I completely agree with you, Jeff. That's really why we in our group pursue long-term therapies. We try to do it with warfarin but it was a failure because people didn't like to use

warfarin. Conceptually I don't think three days of therapy with anything is the magic answer, you know, any drug.

The second thing is with this therapy all we're saying is there's a difference that emerges early and there is no evidence that it is really -- no clear evidence that it actually is lost. If you don't mind, when we get to the combined analysis of the two trials, you will see the numbers being slightly larger and more stable over time.

We also have certain post hoc analysis which, if you'd like, I could show you which looks at higher risk groups. You'll see it is maintained over time. I agree with you that the difference is modest but remember it only emerges after three days of treatment.

As a clinician I want treatments I can use long-term as well. I think combination therapy is the way of the future.

DR. BORER: I just want to clarify one further methodological point.

DR. FLEMING: Jeff, before you leave that point, let me just --

DR. BORER: Oh, I'm sorry.

DR. FLEMING: I think the best -- I think you are raising an issue that for me was also a significant concern and I think the best place to look at this is in the briefing document on page 22 in section 421 where we have the exact data.

The issue that we will be discussing, probably several times later on, is what is the relative importance of a result at seven days versus 35 days versus 180 days. Thirty-five and 180 days were secondary measures in the trial.

We've had extensive discussions about this including in the last meeting that this committee had in October as we were discussing what would be appropriate criteria for trials in this setting. Clearly there was considerable debate about the seven versus 35 days as being the proper primary endpoint.

At a minimum a criterion is that the difference that you see in numerical magnitude must be maintained over time. At seven days, for example, on

death MI we have a difference of 211 events versus 178 events which is in excess of 33 events.

In a relative risk that .83. That relative risk becomes .95. As you go out to 180 days, i.e., a 17 percent reduction becomes a five percent reduction predominately because there are a large number of additional events that were anticipated to not be effective.

But it is of interest that the delta, the excess number of CV deaths and MIs that are prevented that were 33 dropped to 24. That's about 27 percent of those excess deaths and MIs are lost so you're observation was certainly consistent with mine.

When you look at the triple endpoint and adding refractory angina, the relative risk at seven days, .82 goes to .96 so a 16 percent reduction to a four percent reduction or four-fold diminishment in the relative risk. One might say that's because there are a lot of events that weren't expected to be affected longer term.

The excess number of prevented CVs, new MIs, and refractory anginas that were 57 dropped to

29. Not only is it that we're seeing a lot of events that weren't expected to be effected but those that were prevented, 57 numerically, it's in half. It's only 29, by day 180. So I look at these data and concur with you. I see that there is evidence to at least suggest that some of this excess that's prevented at day seven is not sustained longer term.

DR. FISHER: Can I make a comment here? I've heard Tom talk about this before and we disagree in various ways but I think it's important people understand the implications here. I actually agree with the sponsor that things look like they're maintained but I have told them they're lucky because you're adding so much additional noise it wouldn't be surprising actually if there were bigger drops and Tom can do the math as well as I can.

If, in fact, the standard were to be that short-term interventions expected to have effect in an acute setting have to be maintained for six months, and in a setting where most events are going to occur elsewhere, you'll be talking about trials up in the hundreds of thousands. I mean, in order to have

adequate power.

This is a tremendously large issue for drug development. In fact, it would also suggest action to withdraw the approval of a number of additional things also. We can look at it but I think Tom would certainly agree with me and Jeff as well that when you add a lot of additional events, you're going to have quite a bit of variability. That's just the reality of the statistics.

As I mentioned, I actually think it wouldn't have been surprising if the drop had been even more or you could put things together and it can go the other way but there's a huge play of chance as you get out and add a lot of additional noise and I think that is very important to consider.

DR. YUSUF: Actually, I can shed some data on that.

DR. KONSTAM: Can I just -- I mean, I just want to -- yes, Lloyd, in that case why look at six months at all and what would you have taken at six months to say, yes, there is a suggestion that the effect is going down as Tom has suggested. Why look

at six months at all based on what you're saying?

DR. FISHER: Well, as I mentioned, this is a very difficult issue because it wouldn't be at all surprising -- and it's important for other sponsors developing compounds in similar situations because they are tremendously at risk in the play of chance.

Even given the fact they think they have the plan they are studying on, their biostatisticians can easily give them the distribution of the changes out par. Certainly I agree. I would be disturbed if it had entirely disappeared, although actually that would also be consistent with chance.

DR. YUSUF: Can I shed some light on this? Can I have slide 57 of my main presentation? These are data from both studies because exactly the opposite happened in OASIS-1. The curves became more exaggerated with time, and this is the point Lloyd says, that the play of chance can be in your favor in one trial or against you in another trial.

You will see here, and I hope this is reassuring, this is what the two trials together, the totality of the evidence on CV death/MI is. It's

really no hint of a loss of that.

Jeff, we're doing a analysis of all the thrombin inhibitor trials. At least up to 35 days we have been able to get the data from the big trials and there is absolutely no difference, no loss in the benefit that you see at 72 hours. It's just parallel. If this is a crucial issue, we can give it to the agency.

DR. BORER: Okay. I have one additional question. I just want to make the point here and it doesn't require a discussion and it doesn't require a defense.

I agree that a three-day treatment shouldn't necessarily be shown to have a persistent benefit six months later by itself but if it doesn't and the goal is to prolong life or prevent major events and you can only do it over three days or four days or five days, then it seems to me it's incumbent upon those who would propound the use of that treatment to show what the follow-on treatment is that does maintain the benefit. Otherwise, you are exposing people to risk for no apparent benefit.

Again, that's in the record now and it doesn't require a discussion.

DR. KOCH: I think the only further comment that could be added to that is that because as the rates increase, the variability increases and you don't have statistical power at those later times. All you can show are favorable trends. You cannot demonstrate significance in the usual way.

DR. BORER: One final question here if I can find it. I think you may have answered this already but I want to know the differences in the 63 patients that were excluded from the ITT for the MITT assessment. If I understood correctly, you actually answered that question earlier in that most of these people never got drugs. Can you say it again?

DR. YUSUF: Except for three. I think three people were lost to follow up at that stage. Is that what it is? Okay, Gary, you have the numbers.

DR. KOCH: M-105. This is the display showing the patients that are not in the MITT. It's also showing the patients who were crossed over. That is, 10 patients originally assigned to lepirudin.

Sorry, 10 patients who got lepirudin should have gotten heparin and five patients who got heparin should have gotten lepirudin. None of the 30 patients at the bottom had events so the patients who had a misrandomization did not contribute either way.

M-106. These are the patients who were excluded from the MITT because they never received study drug. There were two patients who didn't have a seven-day endpoint and neither of them had events. Among the 63 excluded all together, there were two patients with CVD or MI added to the heparin group or added to the lepirudin group. Basically these are pretty much what you would expect by chance. The exclusion from the ITT is a chance type event.

DR. BORER: Thank you. Milton, I'll hold my other questions. They are not on primary data clarification.

DR. PACKER: Ileana.

DR. PIÑA: Salim, you seem to have abandoned the severe angina as an endpoint here. And then at the other endpoint you pooled out the revascularization issue. Why did you do that?

DR. YUSUF: I think we were going for clinically more important endpoints in the big study. Second, being an international trial, severe angina was going to be a nightmare to adjudicate. We actually found that because we had to get every episode of recurrent angina and it would have just drowned us in work.

We wanted to go for the more clinically relevant endpoints. The refractory angina we felt was pretty objective. We could document it. We could verify it. CV death/MI obviously was that so that was one. Interventions obviously has health care resource implications so we wanted to find out information on that.

It was also mentioned in the protocol for OASIS-1 but we knew the event rates would be so low there wasn't going to be much in it. I'll show you some data on it as well.

DR. PIÑA: Don't you think that is very geographically mediated? In other words, there are some countries that will do this very often and some that just won't?

DR. YUSUF: You're absolute right. There are geographic variations so the analysis was stratified by center because obviously a center with a cath lab will do it more often than a center without a cath lab. The analyses on those -- in fact, all our analyses are stratified by center which takes into account region as well.

DR. PACKER: Paul.

DR. ARMSTRONG: Can I just follow up? Salim, the definition of refractory ischemia was different in the OASIS-2 than the OASIS-1 study. How did the new definition affect the frequency of refractory ischemia as previously defined in OASIS-1?

DR. YUSUF: It would have a minuscule effect and the reason for that is when we did OASIS-2 because of the fact that practice patterns in different countries were different in terms of discharging people earlier out of the hospital. Based on our registry we had found Australia was the one that was discharging people the earliest, U.S. the next, and then the other countries kept them in the hospital.

We needed a mechanism of capturing events all up to the same day. That's why that was brought in so OASIS-2 actually has an added criteria compared to OASIS-1. I showed you the rehospitalizations for unstable angina on one of the previous slides. I don't know if you recall that. That is the difference.

As you can see, they got rehospitalized within seven days so this is being discharged at three days to being rehospitalized before seven days. That's the data. There was a difference but the entire difference was in those who were rehospitalized admitted to a CCU, had EGC changes, and had an intervention.

DR. ARMSTRONG: So when you used the new definition the event frequency of refractory ischemia in OASIS-1 went down or up? I'm just not clear.

DR. YUSUF: We don't have a way of using it in OASIS-1 because, for one thing, I think patients were by in large discharged around seven days in Canada. Not all but by in large. We are not able to retrospectively go back and do that analysis.

DR. PACKER: Jeff and then Joann.

DR. BORER: I'm sorry. Just one more methodological question or analysis question, I guess. I asked you and you answered what the nonstudy heparin issue was in OASIS-1, but nonstudy heparin also was given in OASIS-2 and there are larger numbers here.

Do you have an analysis that would tell us what the potential or actual effects of nonstudy heparin -- I don't care about heparin that was given before so much as heparin that was given after the three days of therapy, what impact that may have had on outcomes with and without lepirudin.

DR. LUZ: Can I have 042? This slide shows you the influence of nonstudy heparin on both the double and triple composite endpoints in seven days. There are several observations that one can make. First is that the use of nonstudy heparin within 24 hours after end of the infusion was higher in the heparin group than was in the lepirudin group.

Second, and this is not unexpected, the event rates for both the double and the triple endpoint were higher in patients using nonstudy

heparin after the end of the infusion because in many cases those patients had ongoing symptoms and this was actually the reason why they were put on heparin.

Second, the difference between the two groups, if anything, was slightly bigger, at least in absolute terms, than in the overall population. Importantly, the entire difference was accounted for by the treatment period, i.e., the period during which the active study medication was given.

If you look at this slide, you'll see that in the time between end of study infusion in seven days, the difference is actually very minor. The same is seen with the triple endpoint where you have a full percent absolute difference in all patients using nonstudy heparin but only one percent difference after the end of the study infusion, i.e., while nonstudy heparin was used.

DR. YUSUF: I think, Jeff, this is an indication that the symptomatic benefits of lepirudin persisted for a few days versus unfractionated heparin so it's 900 more -- 100 more in the unfractionated heparin group compared to lepirudin in that 900 versus

800 is highly significant but we're not making much of that because we really didn't specify that as anything. The event rates were different.

DR. BORER: Thank you.

DR. PACKER: Joann.

DR. LINDENFELD: I'll save my question until we get into safety.

DR. PACKER: Okay. Tom.

DR. FLEMING: In the adjudication process there were 25 MIs, is that correct, that originally had been identified by the investigators and were not confirmed?

DR. YUSUF: You're ahead of me, Tom. Let me see. Do we have a slide on the adjudication? This is 1 or 2?

DR. FLEMING: I think it was OASIS-2.

DR. YUSUF: OASIS-2. Do we have a slide on the adjudication in OASIS-2? Okay. Here it is, Tom. This is the confirmation rates. It's not the absolute numbers. It's the rates. You'll see cardiovascular death was 100 percent, new MI was 95 percent, 98 percent refractory angina. These are for

rehospitalization with ECG changes or without ECG changes.

Do we have another slide with the relative risk adjudicated and non-adjudicated? We have one. I know. Tom, here are the data for CV death and MI at 72 hours and seven days, adjudicated and the investigator reported. You will see the relative risks are the same. The p-values are almost identical.

DR. FLEMING: Can you go back one slide?

DR. YUSUF: Sure.

DR. FLEMING: So five percent of the new MIs and two percent of the new MIs were not confirmed. I think what I had seen in the briefing documents was that there were 25, eight on Refludan and 17 on heparin. This seems to be consistent with that, about twice as many on heparin.

DR. YUSUF: That's right. So if you use the investigative report that the differences would slightly widen out.

DR. FLEMING: If we use the investigators, we would add back eight events on Refludan and 17 on

heparin.

DR. YUSUF: On heparin. That's right.

DR. FLEMING: And if that's at seven days and if there is an excess of 33 in the adjudication analysis, then wouldn't that difference of 33 excess drop to 24?

DR. YUSUF: No, it will increase 33 plus 11. You have to add back. It will go the other direction, Tom.

DR. FLEMING: Go ahead and show the second slide.

DR. YUSUF: The next one. You see, you have one percent more events in the unfractionated heparin group in the investigative report. Not one percent but .1 percent. 2.6 would go up to 2.7 and two remains the same. Remember there's a little bit of rounding here. Then for seven days this 4.2 goes to 4.3, 3.5 goes to 3.6.

DR. FLEMING: And the relative risks?

DR. YUSUF: Are essentially identical.

DR. FLEMING: Slightly less but essentially identical with the investigator.

DR. YUSUF: Yes.

DR. FLEMING: All right. Could we see the same thing for the triple endpoint, though? That looks like it went --

DR. KOCH: Tom, whenever you maintain the same difference if the rates increase, the relative risks will also always mathematically increase towards one.

DR. YUSUF: So these are the triple endpoints, Tom, and this is the adjudicated on top and the investigative report. You will see again they are essentially the same. Obviously relative risk is sharpened here because equal numbers of people are being thrown out in both groups. Well, similar numbers. Equal is an exaggerated term but similar.

In both cases all the analyses are nominally statistically significant. In fact, in OASIS-1 it happened the other way around. Adjudication brought the relative risk up rather than down so I think it's a random process.

DR. KONSTAM: Just to clarify, if you want back two slides in terms of the triple endpoint, there

were more -- if you look at the 98 percent versus the 91 percent and there presumably that's an endpoint with a greater degree of subjectivity, there were more such events thrown out in the hirudin group than in the heparin group.

DR. YUSUF: Slight differences.

DR. KONSTAM: Okay.

DR. YUSUF: These are the data, Marvin, and it was done blind. One can obviously argue they are not significantly different but these are the data and these are not -- all I can show you is the data.

DR. KONSTAM: So the only even really noticeable difference is the triple endpoint. The significance level goes from 01 to 02.

DR. YUSUF: To 02. They are consistent. I think that's the point.

DR. FLEMING: There was an algorithm in place that if the seven-day form was missing that you would count the events as deaths or MIs if it was known to have occurred according to the FDA summary on page 13 of our briefing document. How often was the seven-day form missing?

DR. YUSUF: Seven in each group.

DR. FLEMING: Seven in each group.

DR. YUSUF: At the seventh day. We could have had it on the sixth day. You see what I mean? To be honest, we didn't put that algorithm in the investigation. It was something the company had put in its plan.

DR. FLEMING: In those cases were there any situations in which death or MI or had occurred?

DR. YUSUF: I don't know the answer to that. Is Janice here? Okay. There wasn't any.

DR. FLEMING: My last comment really is kind of a two-fold part. It does lead somewhat into the concepts of the pulling of the data. One of the issues that is problematic is looking at the clinical relevance of the effects of an intervention where you expect the influence on the endpoints to be early, to be in the first three days.

Yet, if in fact those differences that occur in the first few days are not sustained for a clinically reasonable period of time in any setting, we would discount those as being important. As we

have already begun the discussion at this session, it does impact how we interpret the primary endpoint and whether it should be day seven, day 35.

What's the relevance of interpreting day 180 if, in fact, we were to choose day seven as the primary endpoint and information on day 35 or day 180 is just used to show that you sustained the benefit.

As Dr. Fisher and others have pointed out, myself and many others have discussed in previous sessions the down side of the sponsor. The risk the sponsor takes there is with considerable variability that exist. In those events that occur after the time you have an effect, you may randomly miss the event. You may randomly miss an effect that is sustained.

Of course, the flip side to that is it's an effect that is very small in the context of what is really clinically important to these patients. If you're looking at death and MI, and death and MI is an event that's going to occur in 10 percent of these patients, and only a small fraction of those are occurring in the period when the intervention has an effect, then it is true that you're going to have to

have a very large trial.

But, then again, you need a very large trial in order to be able to discern the difference between an effect that is real and sustained versus one that doesn't, in fact, have persistence. It's led many of us to say for this reason you shouldn't, in fact, rely on seven days hoping to see it sustained because by chance you may be unlucky and you do, in fact, need to be looking at differences of larger magnitude or over a longer period of time.

That leads me to my comment looking at OASIS-1 versus OASIS-2. It is true that OASIS-1 tends to show the opposite, i.e., instead of losing eight or nine deaths or MIs you'll pick them up. In a sense if you do a meta-analysis, everything will look fine.

Actually, I'm more concerned by OASIS-1 data than reassured by OASIS-1 data. The specific reason for this is that there is, at least by appearance, a striking inconsistency between OASIS-1 and OASIS-2.

I'm leading up to a question, Salim. The primary analysis that we have been drawing attention

to in OASIS-1 was the four component endpoint at day seven which is the only one that in a sense by rigorous statistical methodology might be called significant. This corresponded to a relative risk of .56 or a 44 percent reduction in the event rate comparing control against the .4 dose regimen.

This 44 percent reduction in OASIS-2 on the same endpoint was only a 10 percent reduction. The relative risk was .9. Let me finish. This, in fact, was seen across the board on other measures. If you look at death and MI, the relative risk in OASIS-1 is .53. The relative risk in OASIS-2 is .83.

The fact that the differences that there are, in fact, in death MIs and increase in the excess that's prevented between day 7 and day 180 whereas there's a decrease in OASIS-2 makes OASIS-1 and OASIS-2 even more inconsistent when you look longer term.

The relative risk in OASIS-2 at 180 days is .95 on death MIs, whereas it was .76 in OASIS-1. The exact same thing happens when you look at the triple endpoint as well. There is anywhere from a three to five-fold larger effect in OASIS-1 than in

OASIS-2.

Looking at OASIS-1 it's not surprising that one might have projected that it could have been plausible to achieve superiority. I think you were targeting a 23 percent reduction in the sample-size calculations. In fact, that was reasonable. Looking at the OASIS-1 magnitude of effect it was even larger than that, whereas the OASIS-2 magnitude of effect was much smaller.

I'm not reassured in this loss of effect by the fact that it's not showing up in OASIS-1 precisely because that's just adding to the ways in which OASIS-1 effects are strikingly different than OASIS-2. Is there an explanation for that?

DR. YUSUF: First, I don't think the smaller the trial, as we all agree, the point estimate of the relative risk reduction is weaker. Therefore, you know, all of us over years have tended to think of the confidence intervals around so when you look at the confidence intervals of OASIS-1 and OASIS-2, there is no heterogeneity in the results at any time point. There's no statistical heterogeneity at any time

point.

The second thing, Tom, we all know that small trials, the best you can get out of it, even if you can get that out of it, is directionality of effect. That's why we do bigger trials. You're right. In the large trial we expected a smaller reduction. We designed it around that.

As you know, I have written for years articles saying that the real effects are going to be moderate. In my mind I did not expect a 40 or 50 percent risk reduction so whatever I saw in OASIS-2 had wide confidence limits. It was consistent with an effect at about 20 percent.

All I'm interested from OASIS-1 is that it gave me a direction that was promising enough to go to the next step. Obviously, if you get a big difference early and there's no loss of that big difference early, you'll see a bigger difference late. In some ways I completely agree with you. Take a small trial but don't believe large treatment effects. Believe it best the directionality of the effect.

Now, the best estimate of treatment from

a bunch of trials is not choosing the most effective difference or the smallest difference. It's to take a weighted average of all those. As you know, that has been my approach for 20 odd years and that's what people have been doing.

If you take the aspirin trials, you'll get trials where there's a 50 percent risk reduction. You'll get other trials like the Amos trial with aspirin which had a zero percent risk reduction. We all accept that the best estimate of the effect of aspirin is the totality of the data. The data from larger trials will contribute more to that totality than from the small trials for they all contribute.

DR. FLEMING: I'm not persuaded by a statistical argument that this is purely random variability. Is there any other factor that you're aware of in the two trials that would explain what is from the estimates striking differences?

DR. YUSUF: We've looked for biases, as I've told you, and we've really looked as hard as we can and we have no evidence of bias. I really believe it's random noise. In the discussion part of the

OASIS-1 paper, we said random play of chance may have exaggerated the differences so that when a larger trial is designed, we should look for more moderate benefits.

That's why when we went to the company we said we really need a much bigger trial, not another trial of using the estimates of effect size. And, Thomas, you know that to be completely consistent with the way I've done trials over the years.

DR. PACKER: Rob.

DR. CALIFF: In light of yesterday I want to ask two quick questions. First, the geographical distribution you label as North America but not U.S. Two questions related to that. First, how many patients were enrolled in the U.S? And, second, how many African American patients and what were the results of those patients?

DR. YUSUF: I think there were about 350 patients from the U.S. in OASIS-2. As you know, OASIS-1 was entirely done in Canada. How many African Americans did we have, 60 or 80? Eight in OASIS-2.

DR. CALIFF: Eighteen?

DR. YUSUF: Eighty, eight zero. In OASIS-1 obviously we may have had one or two Blacks but certainly not African Americans but African Canadians.

DR. CALIFF: Again, I just want to make a point. This is a real problem and I think it almost seems like in the panels I'm on now the majority of studies have almost no representation of black patients from around the world, much less --

DR. YUSUF: You know, what surprised us, Rob, was you know we did the study in Brazil as well as in South Africa and we just weren't getting Black patients in because, you know, I have an interest in ethnic variation in disease and I try to use my trials to study that but we just weren't getting them and the reason is at least in South Africa blacks have not yet got the epidemic of cardiovascular disease and I think the U.S. is leading the way amongst individuals of Black origin in getting the cardiovascular disease. Right now as an inside we're doing studies in Africa and we're told that in the capital of Botswana, Gabaron, there's only 12 MIs a year. This just illustrates how hard it is to get African Americans

into trials.

DR. CALIFF: Certainly if products are going to be sold in the United States, we have 15 percent of the population, just to reiterate, that's being left out of the trials and it needs to be taken seriously by the companies that plan to make profits in the United States.

The second question is women. Because of the pharmacokinetics of the drug, you might be concerned about different outcomes in women and men in this case. Do you have a slide that shows the treatment effect in women?

DR. YUSUF: Matthias? I'm sure we do. Subgroup by gender.

DR. KOCH: While that's being gotten, I wanted to respond to Tom's concern about OASIS-1 and OASIS-2. Tom, while OASIS-1 did show a somewhat larger effect on the quadruple endpoint, it turns out on OASIS-2 the effect on the double endpoint was actually bigger in OASIS-2 than it was in OASIS-1. I'm sorry. I take that back. It was on death. I'm sorry.

DR. YUSUF: No, no, no.

DR. FLEMING: In every measure it is strikingly different.

DR. YUSUF: Sure.

DR. FLEMING: Anywhere from a three-fold to a five-fold.

DR. YUSUF: Tom, one --

DR. FLEMING: You didn't volunteer, Salim, and I don't know how much to make of these but I've been trying to think a lot about what is different between those two studies and what could account for them. Some of the things that come to mind are there is a differing heparin regimen in terms of how it's scheduled.

It may be small. There is a difference, though, in the number of people that have increases in heparin dosing. It's 33 percent in OASIS-1 versus 52 percent in OASIS-2. The concomitant meds, as I understand, aren't exactly the same in the two trials. OASIS-1 is not blinded. We've talked a lot about that and whether that could impact and it possibly could.

That was one of the ways in which it could

be showing up. There are some differences and there may be other ones that you are aware of. The data are strongly suggesting that something is different in these two trials and to pull them, which is what we are about to do, is especially of concern when you are pulling results that are so disparate.

DR. YUSUF: Tom, the differences between OASIS-1 and 2 in results are identical to what you would see between trials in any area that when you try to pull -- when we did the thrombolytic trials. The smaller the trials, the greater the heterogeneity results. I really think -- I truly believe, and honestly this is not because I'm standing here, I truly believe it's the play of chance.

I've wrestled about this for years. You get small trials and the point estimate is the least unreliable information from it. The best you can do from that is look for directionality of effects and use the data toward bringing together all the information to get the best point estimate.

DR. FLEMING: Well, the key is and philosophically what you're saying, I think, and we

don't need to get into an extensive statistical discussion about this, but the essence is, and I agree with you, the OASIS-1 trial is a very small screening trial that was done to provide important insights to the design of the OASIS-2 trial.

One of the unfortunately things about it is it gave us a strikingly higher estimate of efficacy leading us to expect the 23 percent reduction, at least, in the sample size calculations of OASIS-2. OASIS-2 has 20 times the data. The essence of the interpretation of efficacy is really driven by the OASIS-2 trial.

DR. FISHER: I wanted to make another point. Tom brings up an issue which is important, but you should remember that under the no hypothesis of no difference if the true odds ratio is one, you can combine all sorts of disparate trials. Where you run into trouble is when you think there is an effect and you look at these different odds ratios and try to understand why there's a difference. And then for the estimate of the effect that people might expect in different clinical settings it's very important.

Under the no hypothesis of no effect in either trial, then as all we statisticians know, I think, it would be appropriate to pull them because you are just trying basically at that stage to say, yes, there's good biological activity.

Tom's concerns are actually more directed towards -- well, forgive me, Tom, because he won't say this, but he is mentally thinking, "Well, gee, if there's an effect here, why does it differ between these populations and this sort of magnitude."

One argument which is perfectly consistent with the data is chance but that doesn't prove it, as was mentioned.

DR. FLEMING: Lloyd, since you're at the mic, you made an important insight at our meetings in August that were coordinated by FDA and the Duke group leading up to the October meeting that we had here of the advisory committee looking at criteria and design guidelines. One of the issues --

DR. FISHER: I now deny whatever I said.

DR. FLEMING: It was intriguing because you had pointed out that if you do a meta-analysis and

you do so, for example, with a random effects model acknowledging the variability that can be existing among studies, you pointed out the paradox that if you have one trial that shows an effect and then you have another trial that shows a strikingly bigger effect but highly disparate in the magnitude from what the first trial showed, that a meta-analysis with a random effects model will actually give you an attenuation or a lower sense of --

DR. FISHER: Well, no. You can even lose everything. I've made the point even more extreme than that. If you have to be able to assess for possible differences between trial, that means one large trial, no matter how convincing, that's just one trial. Throw it out because there's no way of estimating inter-trial variability in the same context.

Most of these trials, of course -- and it could be there's differences because Canada is in one part and the rest of the world is in the other. There's a lot of hypotheses we can come up with.

DR. FLEMING: This, Lloyd, is one of those

situations so essentially we could spend all day but time is limited. My sense is from all of this there is enough that is perplexing about the inconsistency with OASIS-1 and OASIS-2 that -- and since OASIS-1 is only five percent of what the size of OASIS-2 is, OASIS-2 is the essence of the signal that we have.

DR. KOCH: Could we show E-187 just for purposes of completeness on this point? This is the test of heterogeneity across the two studies for both the double and the triple endpoint either using all the doses of lepirudin in OASIS-1 or only focusing on the use of the middle dose in OASIS-1. All four of these p-values are well above .10 so this supports that there is some homogeneity even though there certainly are recognized differences.

DR. FLEMING: Gary, that doesn't support homogeneity. That is so underpowered. It's saying relative to that particular measure, that particular assessment there is statistically convincing evidence of heterogeneity.

DR. KOCH: I understand the point you're making but the underpowered --

DR. FLEMING: Clinically look at the data. There is a three to five-fold difference in the relative risk reduction. That's striking.

DR. KOCH: I understand, Tom, but when you say the thing is underpowered, you need to recognize that OASIS-2 is a very big study so you are looking at OASIS-1 relative to the standard that is expressed by a very large study.

DR. FLEMING: And we all come back to the same point and that is the main signal here is OASIS-2. You have 20-fold as much data and the only reason statistical tests of certain choices don't pick up as being statistically significant is the small sample size in OASIS-1.

DR. LUZ: I agree but one additional comment. OASIS-1 is a purely Canadian study. OASIS-2 was a world-wide study but it had a considerable Canadian proportion. If you compare OASIS-1 with the Canadian population included in OASIS-2, you'll find first that the baseline characteristics match almost perfectly. Second, that the magnitude of effect is, in fact, very comparable between OASIS-1 and this

subset in OASIS-2 with the point estimate for heparin being exactly the same in both studies.

DR. YUSUF: There is a slight modification, like Matthias said, which is the relative risk in OASIS-2 in Canada was slightly better than the rest of the study but it didn't get exactly to that level. Isn't that right? It didn't get to what we had --

DR. FLEMING: I was going to say the OASIS-1 and the OASIS-2 data together next. I was hoping to see the outcomes in women.

DR. DIMARCO: Mr. Chairman, why do we have to see the -- I mean, we've already seen OASIS-1 and we've already seen OASIS-2. I think the committee members can add. Why do we have to see them put together?

DR. YUSUF: I would like to show them to make certain points, John, if you don't mind, please give me five minutes and we're done.

DR. PACKER: Before we go on I just want to see if I clarify this because the statistical issues that have been raised are clearly important.

I must say being a nonstatistician we as a committee all of us who are clinicians struggle with how to weigh the statistical issues.

I guess I must say that I'm looking at this increasingly and feeling like an old country doc when it comes to fundamental principles of drug development. Let me see if I've got this right.

Salim, you did a study called OASIS-1 which was a pilot trial, a trial which for all the reasons that have been mentioned it was considered to be a pilot trial. It was a small study. It didn't really have a primary endpoint. It didn't really have a statistical plan.

It really was an exploratory trial. It allowed you to set up a hypothesis that was attested in a big study, a definitive trial, a blinded trial, a trial with prespecified primary endpoint and assigned alpha, a trial that allows you to specifically address whether a preliminary finding is OASIS-1 was real or not real. Is that correct?

DR. YUSUF: Partly.

DR. PACKER: There cannot be any other way

of thinking about this. The documentation that OASIS-1 was a pilot trial for OASIS-2 is undeniable. You created OASIS-2. You set up the rules. You said the primary endpoint is cardiovascular death and new MI at seven days.

You said that you were going to test that at an alpha of .05 which after corrected for interim wilks is .048. You then pick the secondary endpoint, assign an alpha of .01. I'm not exactly certain how you got all that .01 to spend having spent it on the primary but, nonetheless, you've got a secondary endpoint with an alpha of .01. You needed to achieve a .048 for your primary. You got .0863. You needed to achieve .01 for your secondary. You got .0163.

What are we talking about here? You said you were going to do something. You used the data from OASIS-1 to set up OASIS-2. You did not achieve what you said you were going to be accountable to in OASIS-2. How can you declare victory having lost?

DR. YUSUF: Okay, Milton. This is getting interesting. I think it is how you view evidence and whether you view evidence on a sole isolated p-value

or on coherence of information from the totality of the data. We have been very frank, fair, open in presenting the data. The good thing is we all agree what the data and the p-values are so that's good.

Now we come to the more interesting part. I have believed, and so to a lot of people, that if you hit a certain p-value and you are just above or below it, it's not as if the evidence changes qualitatively. If you are far away from the p-value, sure, but if you're close to it, you are left with a difficult choice. In that circumstances, you look at first internal coherence, you look at the effect on other end points, you look at the effects on those end points in other trials, and then you say does the treatment work or not. I think in the end we are trying to assess whether the treatment works, not whether a given p-value is the thing because the p-values are out there. We all agree what the p-values are. My point is please evaluate this in the context of the totality of the data.

DR. PACKER: But, Salim, if the p-value doesn't really matter, why are you showing us a meta-

analysis?

DR. CALIFF: Milton, let me take up for Salim a little bit here on this one. This is the crux of the issue that really is going to swing the decision. After all, I think what this committee is supposed to do is to help give guidance or recommendations to the FDA about getting medical products available to patients if they are beneficial or withholding them if they are not beneficial or there is enough doubt and we are concerned that there is going to be harm from unleashing this on the public.

There's no question under the way things have been done for the most part in the past the way you described it is exactly the way it's been done. You do Phase II, you generate hypotheses, you test the hypothesis in Phase III. That is nice, from my amateur statistician point of view frequentist view of the world.

I completely agree with Salim that is not a very good way to look at the world but we are sort of -- and the issue that you're raising is whether I

agree or disagree. Can you change the rules post hoc from what you had said you were going to do.

I would argue that the best look at what a treatment will do when it's put out there is a combination of all the data that you have looked at in a reasonable way. I think an effort is being made here to put that point of view forward. It would be a change from the way things have typically been looked at.

DR. PACKER: It's more complicated than that. We have a history and some confidence in the process of determining a decision based on certain a priori rules that we set up for ourselves. We can do that for a single trial. We can do that for a number of trials. It is very hard to know how to do that for a meta-analysis. It's hard to know.

DR. CALIFF: Well, you have more confidence than I do in what we've done in the last five years. I think from this committee there have been several things we have put out there that turned out not to be so good because under the stringent rules we've set up, it might be better to design an

experiment that was less relevant to what the product does when it's out there in the world because you can easily reach this level of proof of principle. I'm arguing there may be another way to look at it that might be reasonable.

DR. KONSTAM: Milton, can I just chime in as well? Salim obviously needs some help. I think the statement that you made, I think we really need to defer, you know, on judgment here. I think if you really want to get into it, they are not asking for an indication after all that the drug is better than heparin. I think that --

DR. PACKER: Marv, no, no. Please, we need to be careful because there is a sequence of presentations today which are very important. The sponsor is concluding in their briefing document and in their presentation that the combined data from OASIS-1 and OASIS-2 provides persuasive evidence for the superiority of lepirudin versus heparin.

DR. KONSTAM: Well, I didn't see that wording in the indication so --

DR. FISHER: I was going to say maybe we

ought to wait because this is an active-control trial and we seem to be forgetting that here. I guess I won't go into the carbatalol story.

DR. KONSTAM: I didn't see anything to that effect --

DR. FISHER: But people have argued --

DR. PACKER: Why don't we do this. Pause for a moment. Salim, why don't you go on with your next few slides.

DR. FISHER: Just one other fact that Salim will soon say. This combination was actually suggested by the agency. It didn't come from --

DR. PACKER: Wait a minute. Lloyd, what does that mean? Is that a good thing?

DR. FISHER: Well, I'm sure within this room we can have a wonderful debate as to whether that's a good thing or a bad thing. Most sponsors will try to bend over backwards to go along with what the agency prefers. For example --

DR. PACKER: Wait a minute. With varying degrees of enthusiasm.

DR. FISHER: That may very well be but,

for example, I believe, and probably Tom Fleming believes for that matter, the modified intent to treat if you have good blinding is preferable because you eliminate noise and you have all the benefits of a randomized study. We discussed all that but we said it's such a minor issue here and it makes no difference. Throw up the ITT so that this will not be a big distraction.

As you'll see in my presentation, I think you can make a fairly strong case on the basis of OASIS-2 alone for approval.

DR. PACKER: Okay. Let's pause.

DR. YUSUF: At least let me present the data and then you can make the judgment.

DR. PACKER: Dr. Talarico. Hold on one second.

DR. TALARICO: There was no recommendation to make a definite combined analysis. We could not find any statement that clearly stated that the two trials could be combined for the analysis.

DR. PACKER: Please let me reassure the sponsor that whether it was their idea or not really

doesn't matter. It doesn't matter. Let's hear the analysis and let's see what we think of it.

DR. CALIFF: Milton, can I see the results in women?

DR. PACKER: Upon the combined analysis or --

DR. CALIFF: I'll accept either one.

DR. PACKER: Why don't we see the combined analysis first and then we can --

DR. CALIFF: Okay. That's fine.

DR. YUSUF: Okay. Can I have the next slide? Now, as Milton said, I actually agree with Milton. Even if the FDA did not say it, I would have done it because I always believe you look at the totality of the data. Indeed, this is what we also did in the manuscript. Apparently there is a paper trail that when meeting with key members of the cardio renal group, not with Dr. Talarico's group. Dr. Talarico is right, there are some minutes somewhere -- I haven't seen it but I'm told there are -- that before completion of patient recruitment and before unblinding, that the two would be looked at together.

DR. FLEMING: Although, Salim, you would say the fact that it was before unblinding, that is, of course, not before unblinding of OASIS-1.

DR. YUSUF: Sure.

DR. FLEMING: So you already knew that you had really positive results.

DR. YUSUF: Yes.

DR. FLEMING: And it wasn't a particularly noble effort not to include it.

DR. YUSUF: Absolutely. Tom, I'm glad you at least say some things come easy to me. The next slide, please.

So these are the results of OASIS-1 and 2 side by side and over all. You will see side by side on the double endpoint of cardiovascular death and MI at seven days directionally there are similar results in the two trials. The event rates are about the same in the unfractionated heparin. As Tom pointed out, the effect size seemed to be somewhat larger.

However, I truly believe the best estimate of the treatment effect is the totality of the evidence and these are the data on the combined

analysis indicating a 19 percent relative risk reduction with a p-value of 0.033 which is nominally significant. Not an overwhelming p-value but it's nominally significant.

These are data on the triple endpoint cardiovascular death, myocardial infarction, and refractory angina. Again, you will see these are the two arms of OASIS-1. If you only included the medium arm, the results are somewhat better but these are the two arms because we are going to the concept of the totality of the evidence and you will see directionally similar results in the two trials. As Tom pointed out, the data will be heavily swayed by OASIS-2 because it is 90 percent of the data. You will see 6.7 percent down to 5.4 percent, a relative risk reduction of .8 with a p-value that is again nominally significant.

The next slide shows you the data on therapeutic cardiac interventions up to seven days, directionally similar effects in OASIS-1 and OASIS-2. And the combined data indicate a 17 percent risk reduction that is again nominally significant.

On the three endpoints across the two trials, at least directionally the results are consistent.

These are the data on looking at the durability of the effects on cardiovascular death and myocardial infarction using the totality of the evidence. Again, you will see it's 2.7 percent down to 2 percent at the end of treatment which is a .7 percent difference which remains essentially unchanged right throughout the trial so that the only difference on the totality of the data is persistent right throughout.

The next slide shows you data on non-cardiovascular deaths. There were no non-cardiovascular deaths in the lepirudin group at seven days, but after that and by the end of the study, 34 such deaths in the unfractionated heparin group and 22 in the lepirudin group.

If you add this to all the others, we would show a slight divergence but I believe that would be inappropriate because I have no reason to believe hirudin or unfractionated heparin will have a beneficial effect on these endpoints. Therefore, I

believe the combined CV death MI is the right one to use, although this would exaggerate the p-value a teensy weensy bit.

Therefore, the overall data from OASIS-1 and 2, the entire program, I believe, provides persuasive evidence that lepirudin is superior to unfractionated heparin in this population. Remember, yesterday I used the word convincing. Today I'm using the word persuasive.

Now, this is based on the following, that cardiovascular death and MI and the triple endpoint are significantly reduced at 72 hours and at seven days. There is also an additional reduction in therapeutic cardiac interventions at seven days.

There is at least directional consistency of results from OASIS-1 and 2 and no evidence of statistical heterogeneity on the results, although a strong point about that doesn't say we have homogenesis results. The absolute benefits that we observed early are preserved long term.

I think I should now say thank you very much.

DR. PACKER: Okay. Marv, I just wanted to clarify the fact that both the sponsored document and the sponsors presentation, in fact, says either clearly superior or persuasive evidence for superiority. Let me just make a point.

The reason for asking the issue is because if one concluded that this agent was superior to heparin, there would be no need for Dr. Fisher's presentation.

If this committee believes in the conclusion stated in the last slide, then there would be -- we could shorten this meeting considerably. That was the only purpose for my asking the question.

DR. KONSTAM: But nobody believes that.

DR. PACKER: We will find out. We will go to now questions on the meta-analysis. We'll begin with Jeff.

DR. BORER: Actually, I have no questions about the analysis itself. I think the issue is the appropriateness of the combination and I think that is really for a later point in the discussion here. I think you've presented the data and I have no more

questions about the data.

DR. PACKER: I am a little bit remiss. Dr. Califf really would like to see the data in women.

DR. BORER: That's a question.

DR. PACKER: Do we have the data in women anywhere?

DR. YUSUF: These are the data, Rob. This is younger and older, males and females directionally similar results. These are the data by race. These are the data by weight and weight is an important issue as will be discussed later on the safety part.

DR. CALIFF: This is somewhat to make a point that I think particularly with drugs that are renally clear, we need to look specifically at women. We are learning that about QT interval prolonging drugs and I think with any thrombotics that are renally clear because of that intersection of weight, creatinine, and gender.

DR. YUSUF: And we'll come to that in the safety because that's your concern. Isn't it?

DR. CALIFF: Yes.

DR. PACKER: Okay. Ileana, I think you

were next.

DR. PIÑA: Salim, in your combination here of both trials, you have therapeutic cardiac interventions up to seven days. Why did you include them when that alone was never an endpoint in OASIS-1 except as included in the definition of refractory angina?

DR. YUSUF: Why did we include that in OASIS-2?

DR. PIÑA: No, no. I'm saying why do you use this analysis in the combination as if it had been an analysis separate in OASIS-1 when, in fact, it wasn't?

DR. YUSUF: When you do meta-analysis you try to get the same endpoints across all the trials as long as it was collected even if it's not "a prespecified primary or secondary outcome." For instance, many years ago there were at least two instances -- there are three instances I know that this committee did use meta-analysis to assist in the decision making. I was involved in three of them, or two of the three.

The first was the aspirin meta-analysis going back to 1984, in which case no single trial had provided convincing evidence that aspirin post MI reduced CV death or MI. Some trials had mortality as the endpoint. Other trials had CV death as the endpoint. Other trials had MI as the endpoint.

The people who presented it got the same endpoints from all the trials and put it together. When you do a meta-analysis what you're trying to do is totality of information and not any data derived emphasis.

The second thing was when we did the streptokinase meta-analysis, which I think Jeff was the chairman when that came to the committee, there were two bits of data that helped. One was the GC-1 study which was clear on mortality.

The second part was the meta-analysis of 20 trials and there we had some trials that were small that were completely negative suggesting hazard like a two-fold hazard. Others suggesting a 20 to 30 percent benefit but the confidence limits overlap. There, too, some of the small trials, Ileana, did not

have mortality as their primary endpoint.

They had changes in the thrombolytic system or patency but we got the mortality data. The approach of using the same endpoint, even if that endpoint was not a primary or secondary endpoint, is standard in meta-analysis.

DR. PIÑA: In OASIS-1 did you collect cardiac intervention separate from refractory angina?

DR. YUSUF: Yes, we did.

DR. PIÑA: In other words, an investigator may have decided to do an intervention even in the absence based perhaps on cardiac cath data. You had collected those.

DR. YUSUF: Yes, systematically it was collected on a standardized form.

DR. PACKER: Rob.

DR. CALIFF: I don't think anyone is going to argue with the numbers that you've aggregated here. The numbers add up and I guess I've already said that in general I would be in favor more of development programs that stated out front we're going to combine all of our data because it seems to me that there is

some risk and I'm sure it's happened to this committee before that you see the Phase III trial which comes in just under the wire but you don't see all the other data which may not be as favorable. I guess the worrisome thing, Salim, I know you do have a long history of doing this but would you be up here talking about it if OASIS-1 had tended to make the data go the wrong direction?

DR. YUSUF: That's a good question and, in fact, we wouldn't be here. In fact, I would have tried very hard to persuade the sponsors that the p that was just short of significant really was even more short of significance than what OASIS-2 showed because there is no supported data.

To me the concept of supported data and looking at the totality of the evidence is true no matter what intervention, no matter what the result of one of the trials is. You are absolutely right. Often when we are on the committee we are only presented with the best data. We aren't presented with the all the data.

Let's think of a scenario where we have

two trials done by a sponsor. One hits 0.06 and another one also hits 0.06. Let's say that the difference in area where one hits 0.04 and another one was p .10 totally, I would take the first one as being more persuasive than the second one.

DR. CALIFF: I agree with you but it's just that this wasn't laid out like this in the development program which is the point Milton made and that's the problem.

DR. YUSUF: And Milton is right, and actually Tom made a good point. Remember when we did the OASIS-2 study we knew the results of OASIS-1 so had we even written it down somewhere we would combine the data, you know, in a sense it is based on our knowledge of OASIS-1.

DR. PACKER: Marv.

DR. KONSTAM: You know, I mean, isn't this something analogous to a regression to the mean in the sense that nobody would have done the second trial had the first trial not clearly pointed you in that direction. I mean, to me any attempt retrospectively to combine them is sort of stacking the deck.

I think this is analogous to what Rob is saying. You wouldn't have been doing it. I think if OASIS-2 were more positive than OASIS-1, you clearly would not have bothered us with it. I think it's something like regression to the mean.

DR. YUSUF: I think, Marvin, what you said is true. We are doing a world-wide meta-analysis in all the thrombin inhibitor trials. If you take all the hirudin data including some that were stopped because of "not favorable results."

The point estimates are identical to what we saw and the p-values are about 401. Because of the format that here we only look at the evidence from one agent, obviously we can only give you the two trials with that one agent.

If you take the GUSTO series of trials, the TIMI-9 series of trials and you put it all together, you get the same result and it's there in our manuscript, in the Lancet manuscript. That would probably allay some of your concerns.

DR. KONSTAM: I'm arguing that it doesn't -- you know, that I don't find it particularly valid

to combine the two because --

DR. CALIFF: So, Marvin, you would just ignore all previous data?

DR. KONSTAM: No, I'm talking about this specific situation where you have two studies. One was clearly performed as a pilot to the second and the second would never have taken place if the first had not been positive. To then go back and say, okay, now to make the p-value work let's combine the two just doesn't seem valid to me.

DR. YUSUF: Marvin, we didn't make the p-value work. In fact, if the p-value were wholly nonsignificant, we wouldn't be here. That's right. I think that most of you recognize that my approach to the evaluation of the data has always been on the totality of the data. What I've done here is the totality of the data I have available to me.

DR. KONSTAM: I was just involved with a trial called ELITE-2 which was based on ELITE-1 which seemed like this amazingly positive p-value, and then ELITE-2 which was neutral. I don't think that anybody could view justification to combine the two and say

the right answer is the combination of the two.

DR. YUSUF: Can I just make a point which we may agree on. We don't need to combine it. Just look at them qualitatively side by side. We would agree there is directionality that is similar on the same endpoints at the same time point. The combination is only one step further but visually you would say in both trials the direction is the same. Would we agree on that?

DR. PACKER: There are going to be philosophical issues that are of great interest and importance that will not be resolved today. I think that we have to try to focus on the issues and the data at hand. Let me ask the committee are there any other questions about the combined analysis? Tom.

DR. FLEMING: I have two brief questions. Let me just concur with the statement that we surely do want to look at all the data and that is always critical when an advisory committee is reviewing to know that, in fact, there are other studies beyond these two that are highly relevant to this issue.

We have to be confident we are seeing

everything that's most relevant and we need to look at it in the totality. It's the subtleties in interpretation that are critical and Marv is raising an important one about the regression of the main phenomenon. Where we will have troubles here is when we put p-values on these meta-analyses differences and use that in some way as the basis for a conclusion, as Milt has pointed out, a conclusion of superiority.

I have just two quick questions. At least one of them is really quick. Given the imbalanced randomization, the 433 and the OASIS-1, I take it all of you estimates were based on stratifying by study taking into account the imbalanced randomization?

The second point, and it's what diminishes my confidence that this meta-analysis is telling me something I can really interpret, is the heterogeneity, as I have mentioned, between studies.

Dr. Fisher's insight last August on the difference between a fixed effects and random effects model when you're considering center as random effects, is really getting at what is intuitively of concern to me. When you see such heterogeneity and

the random effects model is particularly sensitive to that heterogeneity across studies. I assume this was a fixed effect.

DR. YUSUF: This is a fixed effect. We've also done it by random effects and obviously the p-values go up a little more, but in every case it's less than 0.05. You're right. I don't hang on the p-value. I mean, I believe in coherence of information and so, you know, just to be reassured, if you do an random effects model, the conclusions still hold.

DR. CALIFF: I have a point. Hanging has two meanings. You said you didn't want to hang on the p-value. It has two meanings.

DR. YUSUF: What's that?

DR. PACKER: It's okay. He'll explain it to you later, Salim. Any other questions or comments? Okay.

What I want to do is take an unusual step because it's very relevant to the process that we are going through today. There is no question to the committee specifically on superiority but it is very important in the thinking process of this committee

and in the relevance of the presentation going forward to make a judgment here.

After discussions with the division, it seemed very appropriate to ask the committee for a vote on the issue of OASIS-1 plus OASIS-2 based on the issue of superiority. The question to the committee is a very relevant question and we will discuss it briefly and then take a vote, a binding vote, on it.

The question is based on the data from OASIS-1 and OASIS-2 individually or combined, depending on your opinion, based on the discussion that has been presented already, do you believe that there is reasonable evidence to support a statement that lepirudin is superior to heparin for the indication being sought?

Let me again ask the question. Do you believe that there is reasonable evidence for a superiority claim for lepirudin over heparin for the indication being sought? If the answer to the committee is yes, we can shorten this meeting considerably. If the answer is no, we will take a break. Okay? Because if the answer is that it is

superior, then the concept of putative placebo is irrelevant. Okay. Any discussion on this matter? Jeff.

DR. BORER: Well, obviously what you're raising here is probably the most fundamental issue that we ever deal with at these meetings which is what are the standards of evidence necessary for approval of the drug. The subhead here is what's the standard for equivalence or superiority versus an active comparator.

The secondary issue there would be the standard of evidence for superiority to a putative placebo all of which are separate but combined issues. What we are being asked is whether the data in aggregate provides sufficient evidence to allow us to say confidently that lepirudin is effective, that it's not less effective and, in fact, it's more effective than a therapy commonly used in the community and that it's acceptably safe for its intended use.

What you're asking us really is to provide an opinion, if we're going to provide an opinion about this, to determine whether there is any objective

standard that we need to apply or whether it's necessary to employ our collective intuition without any a priori objective criteria or do we somehow get in the middle of those two which I'm not sure how you would do.

We also need to know whether it's necessary only to consider short-term benefit, to consider longer-term benefit, to require more safety information before we make this determination, and whether a follow-on strategy needs to be added on. And we need to see an analysis of what happens after everybody gets the follow-in.

There are all these questions. I would have to say that at this point I think there are highly suggestive data that lepirudin may be superior to heparin but in the absence of more safety data, some further consideration of the short-term versus a longer-term benefit issue, I'm concerned about drawing a firm conclusion about this when the objective standards that we have commonly employed don't seem to be met here. I would have to say, although at the end of the day I might vote for approval or something, at

this point I would have to say no in answer to your question.

DR. PACKER: Discussion?

DR. DIMARCO: How can we talk about whether it's superior if we haven't discussed safety yet?

DR. PACKER: Because superiority is in the simplest term here efficacy. Safety is going to --

DR. DIMARCO: But we're not comparing two dose ranging trials where we know we're at the maximum dose and since it looks like safety is proportional to efficacy to some degree, or inversely proportional to efficacy, I don't see why you want to talk about just efficacy until we've looked at the relative safety. If there's a lot more bleeding with a drug and it's a little more effective in producing acute coronary syndromes, it may be superior but no one would use it and so why do we want to talk about it?

DR. PACKER: I guess the question is not so much a ruling on a claim the sponsor is not making or is not asking for. The ruling is an assessment at this point in time whether the combined data from

OASIS-1 or OASIS-2 together or alone allows for a conclusion that one is better than the other on efficacy parameters alone.

The reason is that is not asked of us in any of the questions. Otherwise, I would save it to the end. Normally we would be asked to make that assessment separately for efficacy and safety anyway. It is only in an attempt to move the discussion along in a logical fashion that I'm bringing that point, which is not mentioned later, up in the discussion at this point in time. Any other discussion? Rob.

DR. CALIFF: Well, I guess -- well, this is a tough one for me on this issue because there's no question about the standard of this committee statistically and this doesn't meet it. Even the overview doesn't meet the standard of this committee purely statistically of the past which is .05 squared as we've talked about for primary endpoint.

On the other hand, I do think things need to change to where we're looking at all the evidence rather than just Phase III experience and ignoring the rest. There is a concern that in the tougher areas

that we begin to make development of new therapies almost impossible with the standard that we've had.

We may push the drug development industry into areas that are easier to deal with when this is the number one cause of death and disability in the world and projected to be by even a bigger factor the number one cause of death and disability.

I've been involved in other committees now on several occasions in the last year and we have a tough standard of evidence than other committees. Having said that, by our usual standard this doesn't make it even combining the studies.

DR. KONSTAM: Rob, I mean, it seems to me that your concerns ought to be held off until the end of the day because I think, you know, we're going to get into the fact that the problem is here we've got an act of control.

I think the question that Milton is asking us is simply do the data clearly demonstrate to us that we have a drug that's better than heparin. That is the question. I'm not sure that the answer to that based on rigorous statistical criteria is going to

result in any kind of indictment of the process.

DR. PACKER: Okay. Marv, why don't you begin. Yes or no?

DR. KONSTAM: On this question?

DR. PACKER: On this question.

DR. KONSTAM: We have basically a single pivotal trial with a p-value of .08 on its primary endpoint so I think you have to start there and I don't know how you can really make that any better. I would say no.

DR. GRABOYS: I say yes.

DR. GRINES: I would say yes but not to the primary endpoint. I think again that this is one of these situations in the unstable angina trials where the company may have selected a different endpoint than some of the other drugs we approved. If you look at the secondary endpoint, the triple endpoint, I think it's clearly superior to heparin.

DR. PACKER: Let me just make sure. Tom, you said yes. Cindy, you said yes. Is that correct?

DR. GRINES: Correct.

DR. PACKER: Okay. Tom.

DR. FLEMING: It is true that this assessment certainly needs to look at the benefit and the risk in the context of risk. If we look at what are the measures that the sponsor is focusing on, which is the double endpoint and the triple endpoint, the OASIS-1 trial is not significant on these measures at day 7 or at day 35. The OASIS-2 trial also is not.

There is, in fact, a non-trivial, as we will be discussing later, excess in bleeds and major bleeds. My assessment of benefit to risk, jumping ahead not having thoroughly had the risk data presented to us, is that these data don't establish superiority.

DR. ARMSTRONG: No.

DR. LINDENFELD: No, I don't believe they clearly established superiority.

DR. PACKER: Jeff has voted no. Rob.

DR. CALIFF: A very reluctant no based on our standard.

DR. PIÑA: No.

DR. DIMARCO: No.

DR. PACKER: And my vote is no. That's

one, two, three, four, five, six, seven, eight, nine to two. Okay. Which brings us to the next part of the presentation which we will do after the lunch break. We will take a 45 minute break and reconvene at about 20 minutes after 1:00.

(Whereupon, off the record at 12:38 p.m. for lunch to reconvene at 1:20 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:28 p.m.

DR. PACKER: Can I ask everyone to take their seats, please. It is going to be very important for this -- it is going to be very important for this committee to complete its deliberations in a timely manner today because many members of the committee have flight commitments and we will do everything in our power to retain a quorum of this committee. Some who will be unable to stay until the final votes have already provided me with their votes on the questions that are in front of the committee. That's not so unusual.

DR. FISHER: So I should sit down.

DR. PACKER: No, no, Lloyd. You have most of us here. We'll begin the presentation after the lunch break now with Lloyd Fisher.

DR. FISHER: I was going to open thanking the FDA for making me at home by bringing the Seattle weather but I made the mistake of going out at lunch and there was a strange ball in the sky so I won't make that remark.

You have already heard the primary evidence comparing lepirudin plus aspirin to heparin plus aspirin. However, the U.S. drug regulations are by in large written using the concept that an effective drug should be able to beat a placebo. In situations where placebos are not considered ethical we do the best we can which is to estimate the effect of what might have happened versus placebo.

What we will do here is we will not only use the OASIS data but we will try to put it into context by looking at how heparin plus aspirin tends to compare to aspirin and then combining the two.

Here is an outline of my talk. I'm going to use the odds ratio as a measurement of treatment effect and I will explain to you why I prefer the odds ratio. We'll talk about the selection of the randomized clinical trial data we want to use to estimate the heparin plus aspirin versus aspirin effect.

Then I will put that together in the putative placebo analysis, first for OASIS-2 and then for OASIS-1. Finally I'll combine them with your

permission. If you want, I can skip the third slide. I think the issue will be clear before that. Then I will tell you the conclusions that I have from these data.

Epidemiologists in general and people doing meta-analyses prefer odds ratios to absolute measures of treatment effect for a very good reason and the reason is empirically it tends to be much more stable across different studies. That is, of course, immediately obvious if you had a fixed percentage of -- if you have a fixed odds ratio, then the absolute delta depends very much upon the proportion of the events you observed.

In addition, there are theoretical reasons that suggest why this might be true. I'm not going to mention that in the interest of time but Professor Gary Koch, who is here, and is published on this, the reference on the screen, would be happy to speak to this after the talk if you prefer.

Let's move down to the second bullet here and this is the primary point to be made. The OASIS trials, of course, had only two arms and I'm going to

forget the fact that aspirin is in there because everything will be on top of aspirin, lepirudin and heparin.

Suppose for the moment contrary to the facts that a placebo arm had been ethical and was in the trial. Then it's a simple mathematical identity and easy to demonstrate that the odds ratio for lepirudin to placebo is equal to the product of two terms, lepirudin to heparin odds ratio and the heparin to placebo odds ratio.

Now, from the OASIS trial or trials we have this odds ratio directly from the data you have already looked at and talked about considerably. What we will do now is look at what data are available which might allow us estimate the odds ratio for heparin to a placebo for heparin all on top of aspirin. The idea is the following.

We estimate the first term in this product from the OASIS trials. We estimate the second term from previous controlled randomized clinical trial data. When we have those two estimates, we can estimate the combined effect and put them together.

At least as important as estimating the combined effect is if one is willing to make the assumption, and it is an assumption and I'm sure you'll hear about it from the FDA shortly because I was talking to some of them at lunch, but if you make the assumption that the odds ratio is transportable, if there had been a placebo arm in OASIS that odds ratio would be the same, then we can come up with mathematically appropriate confidence intervals, p-values, and so on.

The selection of the heparin plus aspirin data is described very briefly here. We have a number of backup slides we can get into if you would like, the characteristics of the people in the difference studies and so on. Most of the work was initially done by Oler in a meta-analysis published in 1996 in JAMA. The endpoint that he had that was closest to the endpoint in the OASIS studies was all-cause mortality or new MI. I'm going to use that because that was the endpoint that was uniformly available across the studies. I might also mention I think this endpoint is actually more appropriate

because it's certainly a more meaningful thing to the patient than studying an only-cause specific endpoint.

Because of the length of time that had transpired between Oler's meta-analysis, a MEDLINE search was performed to look for additional randomized studies involved heparin or aspirin or both and really not much had been done, although dalteparin had been compared both to aspirin and also to heparin.

And using the same methodology that I'm going to use to put together the putative placebo data, one can get an estimate of the aspirin versus heparin plus aspirin effect by combining these two studies. You might ask why would one combine these two studies. There's a variety of reasons. One is these studies are the closest in time to the OASIS studies that are being talked about here.

Therefore, one would expect concomitant therapy and so on to be a little more contemporary and relevant. Also, as a general statistical principle, I like to use the maximum amount of information. Finally, I first became aware of this possibility from some other sponsor who presented me with the fact that

these studies existed and they said, Is it appropriate to combine them?"

They didn't tell me what the data was or give me a clue to what was going on. I said yes, and especially in this area where there is so little information it's appropriate to combine them. The reason I bring that up is it will turn out these data give a very favorable, although not out of line, estimate of the heparin effect.

People tend to think of we statisticians as totally objective and computers. Well, I can assure you we're not. A lot of the subtle issues that come up on this committee, it helps to have a standard of the way you behave with respect to things, to know how you analyze the data, what you would do. This was a decision that I made prior to myself being unblinded to the dalteparin data.

One other issue which we've heard discussed today is what level of proof might one expect against a placebo. Dr. Packer claimed that if there was a statically significant difference with heparin, I would submit that wouldn't have been true.

Number one, Rob Califf said people use .05 squared divided by 2 but that's in a placebo controlled setting without a serious irreversible endpoint.

Historically with serious irreversible endpoints one study at .05 has been sufficient. However, because of recent events people would like to see more evidence. As far as I know, there is no agency-wide position on how much evidence. I think it is fair to say that when people come in with a serious irreversible endpoint in one study, that they are often requested to have power for more than a significant level of .05.

This is the first of a series of three build-up slides so let me orient you to this. What we have here are the two sources of data I mentioned; the Oler meta-analysis, which is a combination of six fairly small studies. It was not statistically significant as you can see from the fact that the odds ratio crosses the line at one. That was published and is often sites as the reason the medical community started using heparin and to this day there is a very strong belief, as far as I can see, in the value of

heparin in the setting.

The FRIC and FRISC combined actually. Just this study standing alone has a statistically significant estimated odds ratio for the heparin versus placebo effect and arguably gives you more evidence that heparin is effective and the entire Oler meta-analysis thus justifying what the physicians have been doing anyway. My preferred approach is --

DR. CALIFF: Lloyd, I hate to interrupt you but I want to be sure that we're clear on what FRIC and FRISC were actually comparing. My understanding is one of them is actually low-molecular weight heparin versus placebo.

DR. FISHER: No, the dalteparin is a low-molecular weight heparin and it was compared to unfractionated heparin. The estimate I have come up with -- thank you. This was in my notes and I forgot -- is for unfractionated heparin versus placebo. That's why two studies had to be used.

DR. CALIFF: This is an indirect comparison?

DR. FISHER: Yes, this is indirect

comparison of unfractionated heparin versus placebo. This odds ratio is analog rhythmic scale for a variety of reasons, but perhaps the best reason is that an odds ratio of three is equivalent to an odds ratio of one if you took the treatments in the opposite direction. As I mentioned, here we see evidence that heparin, in fact, is better than placebo when used in addition to aspirin.

What we have here are the data from OASIS-2. We're talking about OASIS-2 up here for the endpoint of all-cause death plus new MI. You can see this is not statistically significant, the p-value .086. That is reflected in the fact that the confidence interval for the odds ratio just crosses 1.

Now what we're going to do is to put the two parts together using not only the point estimates from the two parts but also the estimates of the variability from the two parts.

This is my preferred estimate on the bottom that I will focus on because it uses all of the data. I should mention FRIC plus FRISC looks a little out of line here but, in fact, of the six studies in

the Oler meta-analysis two of the odds ratio estimates were .29 and .36 and this is .35 so it's not especially far out of line of the other data, although a lot of this is based on very small samples.

The other thing I want to mention is on a log scale the confidence intervals are symmetric and you see some things that are not symmetric. That is because we truncated the picture in order to enlarge things at an odds ratio of .25. When you don't see the left-hand side, it's a conservative graphical approach in the following sense. All I have done is thrown out values that are even more in favor of lepirudin than the ones that you see.

Here we have if you use Oler alone, FRIC and FRISC or Oler plus FRIC and FRISC, an estimate granted based upon certain assumptions that had their been a placebo arm in the OASIS-2 trial, that the correct odds ratio, the best estimate is .49 with as confidence interval running from .32 to .75 with the small associated p-value that you have here.

One thing you might notice is one could argue the case that there is adequate evidence for

approval based on OASIS-2 alone just looking at this one slide. That depends upon how robust you think findings have to be and a variety of other factors.

Now what I'm going to do is to show you precisely the same figure both for OASIS-1 and the combination. This is the same data from OASIS-1. The first three lines, of course, are identical because we are using precisely the same control data. In the middle of the slide you'll see the OASIS-1 data which for all-cause mortality and death is just right at statistical significance and the confidence interval just touches a value of 1.

When we combine these you'll notice there's a missing circle on this thing. The reason is this point estimate of .17 is even to the left of .25 so it's been truncated. What you see here is the upper limit of that confidence interval.

Again you can see there's a fairly low point estimate from OASIS-1. As Tom and others have pointed out those data look for favorable and not surprisingly then you have a smaller estimated odds ratio, again small p-values.

This finally is the combined data of -- I guess beauty is in the eye of the beholder -- either appropriately or inappropriately combined. I'll just say it's consistent with the other two slides not surprisingly. Actually, if you go back and look at the other two slides, this looks very much like OASIS-2 alone because, as you know, most of the data comes from OASIS-2. The estimate rather than being .49 is .47 but it is essentially the same.

I have done some other analyses which I will not present unless the panel requires it. Rich Simon has a Bayesian approach using very conservative priors. I have modified that for odds ratios and absolute treatment effects and have that prepared in backup if you would like to see it.

I did an analysis on the percentage of the estimated heparin effect preserved by lepirudin. All the point estimates, of course, are greater than 100 percent since things are close to statistical significance. Of a variety of analyses the absolute worse thing I came up with was 95 percent confidence was that the lepirudin preserved at least 88 percent

of the heparin effect.

First, I conclude there is evidence, not as strong as one would like that, in fact, the use of heparin in acute coronary syndrome is not as off the wall as one might argue it could be if you're will to accept these sorts of assumptions. That is statistically significant at the 05 level with FRIC and FRISC alone or with a combination.

Lepirudin plus aspirin looks superior to aspirin with the associated confidence intervals and estimates of the odds ratios and p-values. In my own mind these types of analyses, by the way, have been presented to this committee before. It was presented for clopidogrel compared to the active-control aspirin. A Bayesian analysis was presented for enoxaparin.

I think compared to the enoxaparin the inference here is at least as strong. It's not as strong in some ways as clopidogrel because aspirin had been studied to a very large extent and there were literally tens of studies in almost any area you could think about.

Given that, in my mind I also buttress this by several other things. One, if you use a triple endpoint and you look at the medium dose, then actually both OASIS-1 and OASIS-2 individually are statistically significant. The effect size that we see here, the OASIS data looked very much like other hirudin data from desirudin which Dr. Hirsh will be presenting shortly. Let me stop here and ask for questions and/or comments.

DR. PACKER: We'll start with our primary reviewer, Dr. Borer, and then Dr. Fleming.

DR. FLEMING: Just a quick procedural question just also thinking of efficient use of time. Is the FDA -- we've been given some handouts. Did they intend to provide some comments on these issues?

DR. PACKER: Yes. The sequence that we had envisioned was that we would lead with our primary reviewer, move to our primary statistical reviewer, and then move to the FDA statistical reviewer, and then have a discussion.

DR. FLEMING: In the interest of efficiency to avoid my overlapping what they're going

to say, I'll defer my comments until after they present.

DR. PACKER: Okay. Dr. Borer.

DR. BORER: Yes. I want to preface what I'm going to say with the comment that I really don't think it's appropriate dealing with the issues with which we're dealing to be doctrinaire and rigid and all that kind of stuff. I think that, Lloyd, you made the appropriate point. We have to do the best that we can and there does have to be a way to develop new drugs.

I also have to point out that virtually everything I know about putative placebo calculation I learned from you. In fact, virtually everything I know about analyzing clinical trial data I learned from you.

DR. FISHER: By the way, one other additional point I forgot to add is during the discussions with a consultant from the Cardio-Renal Division, Dr. Lipecky, so I don't pin this on GI, but he suggested -- not surprisingly because the history of this committee -- he suggested that putative

placebo be looked at before OASIS-2 was unblinded.

DR. BORER: Okay. Having said all those things, I need to ask a few things for my edification and probably for everyone else's except for Tom. My understanding of the putative placebo construct that we have used in the past, or that we talked about and use sometimes, is that for optimal confidence one would like to have multiple trials showing quantitatively similar treatment effects of the comparator drug versus placebo. Not just qualitatively similar but quantitatively similar. That's what we'd like to have.

I look at the Oler meta-analysis and it's true that the relative risk varies from .29 to .89. the .29 is based on one event in 69 patients divided into two groups and on and on and on. The point is that what we have here to make up that meta-analysis if you didn't just put the meta-analysis point up there but the individual trials including FRISC and whatever, FRIC and FRISC, is that we have multiple small trials on the Oler meta-analysis with a wide variance. None was individually significant. The

meta-analysis for all of them together wasn't significant at the .05 level.

I don't want to misquote you but I think I learned this from you also at the meeting a couple of years ago about calcium channel blockers that for meta-analyses we generally expect a higher standard of proof; that is, .01 rather than .05 but we'll forget that for the moment. Then we added a different trial using --

DR. FISHER: By the way, Jeff, the standard of proof related to the final end of the thing, not the individual components. If we had a billion small trials, I would be a lot happier obviously.

DR. BORER: Okay. I know you would. I'm just building up to a question here which is your comment. Then we added another trial or pair of trials, FRIC and FRISC, that dealt with a different drug as a comparator for both of our drugs that we wanted to compare so we could compare a third drug to placebo and did some manipulations there. That trial is far more impressive than really any of the

comparators from the Oler meta-analysis and that really drives the significance of the subsequent meta-analysis including FRIC and FRISC.

Then we are judging the effectiveness of lepirudin compared with that commish. Now, I don't say that's wrong and it's probably the best we can do. I may be sufficient to draw a conclusion. But I'm concerned about the confidence that we can have in a putative placebo based on a number of very small trials with a wide variance, etc., etc., etc. I would like you to comment about that. Tell me where I've made my mistake here.

DR. FISHER: Well, No. 1, of course, I would be much happier, as would you be, if before this therapy was widely instituted the medical community had done what to me would have been the appropriate steps.

I guess I also have to say, and I hesitate to say this because one of the quotes I like is, "There's lies, damn lies, in statistics," in my clinical experience." I have somehow gotten the feeling and I even had a line up here about biology

and everybody jumped all over me.

It was perfectly clear to me I couldn't defend myself one biological iota so we took that off. Given the consistency of all these interventions in this area and so on, I guess I have more belief in heparin than the meta-analysis indicates. You have to insert your own judgment there.

I have more discomfort with this than I would had there been a number of larger studies. This is the best that can be done. Actually, in a lot of ways, I'm more comfortable in some ways with the dalteparin, FRIC and FRISC put together than the other studies to be perfectly frank rather than the other way around.

As a general principle, as I've mentioned, unless I have a good reason to throw things out I go with maximum data and certainly there is not as much data here as one would like and I can't manufacture it. All I can do is talk about what's there. I have no magic answer to your concerns.

DR. PACKER: Could we ask the two FDA statistical reviewers to present their review of the

putative placebo issue? I think we'll begin with Dr. Permutt first.

DR. PERMUTT: Thank you. I want to make a few general comments about this method as it is applied to the problem we have in hand. Then Dr. Sue-Jane Wang, who has reviewed it in greater detail, I think has some more specific things to say.

First of all, I heard both Dr. Fisher and Dr. Luz earlier refer to FDA regulations about the standard of approval being a comparison to placebo. Now, I am not an expert in the code of federal regulations and I might be wrong but in connection with this very question, I have looked for that regulation and I can't find it.

I don't think there is any such regulation. What there is in the code of federal regulations is a considerable discussion about the kinds of trials that might be considered to be adequate and well-controlled trials for the purpose of demonstrating what the law requires which is that the drug has the effect it purports or is represented to have.

Now, one kind of trial that is discussed at some length is the historical control trial. I think that is very relevant because that's what we have here. When you are fairly convinced that you know what happens to people if they don't get your drug, then you don't even have to do randomized trials at all because you can do one-armed trials and find out what happens to people when they do get your drug and then compare that to what happened to people who don't get your drug.

That's in the code of federal regulations but it's not used very much and we all know why it's not used. It's because it is very difficult in general to say that what would happen to the patients in your trial if they hadn't got your drug is the same as what happened to patients in the past who haven't gotten your drug. As Dr. Fisher said, the practice of medicine changes. All kinds of things change from one population to another.

What we have here is a version of a historical control trial. We have an unsuccessful -- by its own light according to this protocol and the

committee seemed to agree this morning, unsuccessful trial designed to demonstrate that Refludan is better than heparin.

Quite a large trial, one that if it had produced the expected effects would have been extremely significant. We also have some historical data comparing heparin to placebo in the presence of background therapy of aspirin.

As a member of the committee just remarked, maybe that's the best we can do here. Maybe there is some way of inferring in our minds what the comparison of Refludan to placebo would have been.

But to get to that to a p-value with three or four zeros on it and say you have actually compared Refludan to placebo and you know what the p-value is and so we have the equivalent of a placebo controlled trial so we're done I think is a very big leap.

I think I'll leave it there unless there are questions and let Sue-Jane Wang tell you more.

DR. PACKER: Can we proceed to Dr. Wang's presentation.

(Whereupon, waiting for the machine to

warm up.)

DR. PACKER: Is there any historical data as to how long this takes? Are things in general going in the right direction?

DR. FLEMING: There is a trend.

DR. LINDENFELD: No statistical significance.

DR. PACKER: I have a feeling that the point estimate of delay has wide confidence intervals. When I was in college I used to be a stand-up comic.

DR. PIÑA: Now, that makes sense.

DR. KONSTAM: Now you just sat down?

DR. PIÑA: It finally made sense.

DR. PACKER: It was a long time ago. Maybe you can -

DR. WANG: I think the problem is I really need to show the graphics to explain what's going on.

DR. PACKER: The panel has the graphics.

DR. WANG: So I'll just be talking to the panel.

DR. PACKER: Okay. Let me see. Dr. Wang, we do have the -- I guess it's hard to discuss this in

the absence of being able to point to the slides.

DR. WANG: If the graphics doesn't explain exactly what's going on, when is it appropriate, when is it not appropriate.

DR. PACKER: Does the committee have any other questions they want to ask, Dr. Fisher, while we're waiting?

DR. CALIFF: I certainly -- this FRIC and FRISC routine with no direct comparison between unfractionated heparin and placebo, how much confidence can we have in A is greater than B and B is greater than C. Therefore, A is greater than C by some finite number.

DR. FISHER: Well, number one --

DR. CALIFF: I'm interested in Tom's point of view on this also.

DR. FISHER: Number one, the syllogism that A is equal to B and B is greater than C and A is greater than C is how it turns out. Number two, the confidence went down because there were two studies. The variances were added. But the real concern that the agency has been speaking about is not that part of

it because we statisticians can deal with that. It's the --

DR. CALIFF: Wait. Don't give me this we statistician stuff. That's not acceptable to me.

DR. FISHER: No, no. I was going to point out --

DR. CALIFF: You're telling me if you can deal with it that we never need to directly compare things. We can always confer.

DR. FISHER: If you let me finish my statement, Rob, what I started to say is we statisticians can deal with that but the real issue is the assumption that this odds ratio would have been the same in both studies for the other thing that wasn't there. That is definitely a big assumption.

I'm in print as saying if you can ethically use a placebo, it's unethical not to use a placebo. I don't literally mean that but you really should and I'm no big fan of historical controlled data and everybody who knows me knows that. It's a situation what do we do where you're in an era where people ethically feel that they cannot use a placebo,