didn't understand the third one though. You will have to say that in plainer English.

DR. FLEMING: The third is that if you don't count death as an endpoint and you are looking at a time-to-event analysis on time to MI, those people are still in your analysis post their death. You are imputing, in essence, the time to their MI by the time to MI of other people who were free of MI at the time this person died but didn't die --

DR. LIPICKY: Okay, I retract my statement.

[Laughter]

DR. PACKER: We are making progress! Yes, Tom?

DR. FLEMING: Let me just largely concur with much of the rest of what has been said. As you can tell from this comment, I do believe death should be part of the endpoint. I do believe MI should be part of the endpoint. I have also been intrigued by the PURSUIT data and believe that there is a role for a CEC, but I believe that role of the CEC is, in essence, particularly in an unblinded trial, to get rid of the bias due to unblinding and, secondly, blinded or not, to achieve standardization.

What I think it isn't there for is to change the bar, and I think Rob has given some insights that I didn't have about maybe why the bar was changed so profoundly leading to a doubling of the number of events. It was intriguing there that those additional events that came in

were lesser associated with death, subsequent death, and were lesser impacted by the intervention.

My belief is endpoints ought to be clinically relevant. So, I am swayed to favor those that are investigator-detected. I am encouraged, Rob, by your point that when you had gone back to analyze the data what you found is that essentially those with 2- to 3-fold elevations -- if you had used that instead you would have come in concurrence. My belief on all of this then is that in essence should be investigator-driven assessment tweaked, but not profoundly changed by a standardized overview by the CEC.

The other comment I would make, for reasons we have discussed this morning, is I would prefer this death/MI endpoint to be assessed at 30 days.

DR. DIMARCO: Tom, that only reviews positives, not negatives.

DR. THADANI: No, no.

DR. FLEMING: Not necessarily, you know, you could achieve much of what you wish by reviewing the positives but you could, in fact, look at enzyme-based reviews but, as Rob was pointing out, not triggering any elevation to look at the case but significant elevation to look at the case.

DR. SEIGEL: Another perspective on the issue of investigator-determined versus CEC-determined endpoints is

to the extent that the rates may vary differently, as has happened on at least one occasion, even if the investigator-determined rate is perhaps the better measure of ultimate outcome and more clinically meaningful and, therefore, perhaps even more appropriate in the context of that trial, in the context of the discussion today we should also think about the implications of using the effect size in that trial for planning other trials, whether active control or not.

Let me point out that if the effect size that you are measuring is not one by precept criteria for an MI but by an investigator's judgment, unless you have all the same investigators with all the same judgment in your future trials, given that we see different effect sizes, depending on who calls the MI, it will add another very difficult variable to deal with in determining how to do an active-controlled trial, or even how to size any future trial. So that is an important point of variability that needs to be addressed.

I think on the issue of how different assessments correlate with outcomes, it ought to be noted that at least in some of these trials, I think the trials in question, often the investigator sends in the case report form in which he says whether or not there is an MI not at the time of the MI but after some period of time, such as at the end

of 30 days, at which point of time he may be aware as to whether the patient later died or had arrhythmias and that may influence his decision as to an infarct. So, that is something to think about.

Finally, I do want to get back to that question of inclusion of death. While I agree with everything that Tom said and I think it would be highly problematic to exclude, and unacceptable from my point of view, to exclude death, there is one thing Ray said which I think is important to note, and that is that we shouldn't be requiring a trend in the right direction if, in fact, we expect to have 8 deaths in each arm and if, in fact, we expect a very small effect on mortality or no effect on mortality, then if we really require that mortality trend in the right direction we are going to wind up failing a lot of trials simply because of chance. Half of them will go the wrong way if there is no effect on mortality. So, I would agree with Ray's observation in that regard.

You may want to put some limits on how much of a bad trend in mortality you are willing to accept, or there are other more sophisticated things like a composite can be a rank composite in which death is scored as a worse outcome than MI so that trends in that direction weigh a little more. But to actually require a trend is probably not appropriate.

DR. PACKER: Just to clarify, Ray and Jay, when one uses clinical composites, not just in this area in every single therapeutic area, although we generally say we like things to be concordant across the components, there is considerable leeway in terms of how that is conventionally interpreted, especially if a component is represented infrequently and goes in the wrong direction to a very small degree. I don't think this committee, and I don't think any process that I am aware of has held it against the sponsor, even for an important endpoint like death, when the number of deaths has been small and the numerical difference -- small compared to the number of non-fatal events, and the numerical difference has been very small. Isn't that fair?

DR. SEIGEL: Well, to the extent to what this committee has done, I don't know -- I think that is appropriate, and that is the point I am making, that to actually require -- if you saw 10 treatment deaths and 8 placebo deaths to say, well, that is a failed trial on that basis would be a rather extraordinary and inappropriate thing to do.

DR. LIPICKY: Since I am sort of winning --

[Laughter]

-- the other thing then is that I think the indication shouldn't read "death and" if we really know death just absolutely was not evaluated. I don't mind

including it. But then this thing should not be known for a mortality effect and something else. So, I would be comfortable leaving death in there if I didn't have to put it in the indication.

DR. CALIFF: Now, wait a minute. I went through this when I was a guest on the neurology committee where they approved a combination of drugs for prevention of stroke but not death and stroke. It is the same issue. The problem with that is just what Tom said. You have informative censoring of the deaths.

DR. LIPICKY: Once you see the results, if you really know that deaths are not different, then they can't have influenced your analysis for time to MI.

DR. CALIFF: Yes, they could.

DR. LIPICKY: How?

DR. CALIFF: If the people died in the different groups for different reasons or were at different risk of MI if they hadn't died.

DR. LIPICKY: Oh, I see.

DR. PACKER: I think what Ray is concerned about is not so much the statistical issues but the perceptual issues --

DR. LIPICKY: You don't know you prevented death.

DR. FLEMING: The essence here is death/MI, which does not mean you have proven death alone has been impacted.

That is the understanding we have to have.

DR. LIPICKY: Understanding how the drug becomes known is having influenced both.

[Multi-member discussion]

DR. FLEMING: You have influenced the composite endpoint of MI-free survival, that is true, but that does not mean you have proven an influence on survival itself.

DR. LIPICKY: Correct, but it is known to have done that the way these drugs are currently labeled with that kind of combined endpoint.

DR. SEIGEL: Well, they are labeled accurately, but I think in the sense that they are labeled to have had an effect on a composite endpoint, death and MI --

DR. LIPICKY: You and I understand that, but I think the average doctor doesn't.

DR. THADANI: Also the patient doesn't. The patient thinks you are going to make him live longer when you say death or MI to him. I think that is more implication so --

DR. LIPICKY: Oh, sure.

DR. THADANI: So, what really you are suggesting perhaps is making MI as the primary endpoint and put that as a secondary endpoint.

DR. PACKER: No, no, no.

DR. THADANI: Could you do that?

DR. PACKER: That would create a second disease to treat the first disease.

DR. THADANI: But you are treating the first disease to prevent death.

DR. PACKER: If you wanted to, you could fix this problem by saying that a treatment effect was shown on the combined endpoint of death and MI, and then put a second sentence in "but no effect was seen on mortality alone." If you wanted to fix it, if this is a problem, that would be the solution.

DR. LIPICKY: Well, there are a number of solutions but I wouldn't put the combined endpoint in there.

[Laughter]

DR. KONSTAM: I think it could be easier than that because, you know, I think that what people are saying -- what Rob is saying, and I think everybody should agree with this, is that it could happen that there are deaths occurring that are preventing -- that if your primary endpoint was just MI, there could be deaths occurring that are preventing people who would have been having MIs from having MIs, and you don't want to have that happen, and so if you say, well, MI is where the money is and that would be my endpoint but you are going to have to look at the composite of deaths or MI for those reasons that, de facto, becomes the endpoint.

But what Ray is saying is when you start wording it that way, and I was on the same panel with you and I think we agreed and still wound up saying different things, which is common, because I don't know that there is a way of wording it that gets you away from the concept that you have, in fact, impacted on death.

So, maybe you just say when you word it, and this is what Ray is saying, you say the effect is on MI but we accept the fact that when we count MIs we cannot construct an endpoint that does not include death in the endpoint.

DR. PACKER: I am going to have three other people comment on this and then we are going to go on to the next presentation, regardless of any other points raised by their comments. Bob Fenischel first?

DR. FENISCHEL: Yes, I think that this in part repeats what Marvin just said, which is that the notion of what goes into the label and what constitutes a statistically defensible set of endpoints for use in defining a trial may be entirely different. What goes into a label is the summary of what we think happened in the body of trials and everything else we know about the drug, and a little bit of a hint about what we think we know about sister drugs, and all sorts of other information that somehow all pulls together.

I think in the past few years there has been a

tendency, reinforced fairly heavily by the committee, to say that labeling ought to reflect in exquisite detail what happened in the trials, as opposed to reflecting what we think we understand from the trials. I don't see that there is any contradiction in the law, certainly, to saying the trials must be constructed as Tom has instructed us, which is to say a composite endpoint has to include miscellaneous occurrences which might cause informative censoring, and then we step back and say what we know about the drug; what is the drug supposed to do; what may the drug be expected to do. Well, what we see from the trials is the drug prevents MIs.

DR. PACKER: David, you had a comment?

DR. KONG: Two issues. One is you can reflect the labeling for combined MI and death as this drug will reduce the likelihood of MI but not kill you while doing so. Number two is --

DR. PACKER: You can't get there from here because your confidence intervals go from here to New York.

DR. KONG: Right, but, you know, the idea being that you can demonstrate that you don't have a significant adverse effect on mortality --

DR. THADANI: You can't say that. The sample size is too small to say that. You need thousands and thousands of patients.

DR. KONG: Right, and that is why we can't use mortality alone. Number two is with respect to determining endpoints from investigator-adjudicated decisions, in the setting of active-controlled trials one of the potential biases if you are using non-inferiority arguments is that the investigators could potential bias the outcome by simply saying, "the more outcomes we report, then the more likely the treatments are going to look like each other," and that would be a reason for using central adjudication perhaps even blinded to whether you are doing a superiority or non-inferiority analysis.

DR. PACKER: Jeff, you have the last word.

DR. BORER: I would like to propose a principle here, which is that the reason to develop drugs and to approve them is to reduce the burden of disease on a patient as opposed to just reduce death. If that is true, then, you know, we have talked about death as one of the endpoints, MI as one of the endpoints, but I want to come back to the refractory ischemic symptoms because the existence of these is a situation that is not tolerable for a patient, and it is not unreasonable for a drug to minimize that burden.

Now, I think Rob's point is a very good one, as usual, that is, if you add that you don't know whether the power of your study is going to be improved by adding more endpoints or reduced. We don't know. It depends on the

pathophysiology underlying the problem and the action of the drug. But I want to ask, and I would like to hear Tom's comment even though Milton may not allow him to do it -- Tom's comment about the impact of eliminating the refractory symptoms. What do you do with them? Do you censor a patient at that point? If so, we know that people with unstable angina actually are more likely to have an MI or die than people who don't have unstable angina. So, you are clearly informative censoring. If you leave them in but forget about the fact that they had unstable ischemic syndrome and say, well, did they have an MI or death after they had a PTCA or after they had a bypass, doesn't that confound the data in some unusual and totally unpredictable way, and probably adds events, and perhaps adds events in a biased way? I mean, how would you deal with it if you forget about the fact that people have refractory ischemic syndromes?

DR. FLEMING: If we viewed that those events were of comparable clinical importance to death and MI for the reasons you pointed out, it provides in essence a cleaner endpoint to include them. On the other hand, if they are not and, as a result, it really alters the interpretation of the endpoint I would continue to favor death/MI. As you point out, the occurrence of that endpoint, death/MI could be influenced by supportive or concomitant therapies. In my view, that doesn't require that those concomitant or

ancillary therapies, when they occur, have to be factored into your endpoint. Generally, I would argue, they should be described so my believe is I would still follow for death/MI. If someone had an urgent revascularization I am still following that person to the endpoint of death/MI if they hadn't yet had an MI at that point. I would then describe in my report, in addition to the effects of intervention on the primary endpoint, whether there was a differential experience with these other interventions.

DR. PACKER: Thank you, Rick, very much. Let us proceed to Keaven Anderson's presentation. Everyone will be reassured that we are right on schedule.

[Laughter]

Timing of Endpoint Analyses

DR. ANDERSON: Thank you.

[Slide]

I am going to talk in practical terms somewhat about what our recommendations would be using trial results, talking basically about trial results supporting abciximab as an active control.

[Slide]

There are basically two points I want to try to make through the talk, and I will focus mainly on the second one. First of all, based largely on the discussion that has already been reviewed here today, we think that there is no

really appropriate active control established for ACS trials without PCI, and that abciximab is the appropriate active control for PCI trials.

[Slide]

So, very briefly, in trials without PCI, first of all, with tirofiban in PRISM and PRISM-PLUS the short-term but not longer-term death and MI benefit shown in PRISM and, obviously, PCI was encouraged to some extent, or at least cath, and if PCI was performed it was encouraged during the course of the study agent. Eptifibatide, PURSUIT, there is a small absolute benefit in the 30-day primary endpoint when you have mixed the population as they have. Michael certainly discussed the advantages and disadvantages of a large simple versus more direct trials. But, in any case, it was noted that with different strategies and different subgroups within consistent results it would become possible in active-controlled trial to rig it in some sense so that you can more easily get a positive result in your trial.

Along these lines, we are currently doing a medical therapy trial essentially where PCI is discourage during the study drug infusion. It has a placebo control and a 30-day time point of death and MI.

[Slide]

The one trial that we have done where there is a little bit of information on medical therapy for abciximab

is the CAPTURE trial. These were patients who had refractory unstable angina, and also had had an angiogram and were planned to have PTCA, but there was a medical therapy period of 18-26 hours prior to PTCA.

What you see at the 30-day primary endpoint was that death, MI and urgent intervention was reduced from about 16 percent to about 11 percent, about a 5 percent absolute difference, and this was in 1165 patients. So, it is over 500 patients per arm.

[Slide]

This kind of classic slide from CAPTURE is looking at incidence of myocardial infarction before and after PTCA. On the left-hand side we censored patients when they went to PTCA. In this case, it was mainly because PTCA was planned. Also, in the placebo group, more often when they went to PTCA it was due to urgent symptoms. So, the differences may have been even greater even though the censoring could be partially informative.

In any case, there is a fairly small event rate during the medical therapy period. Then, actually we reset time zero to the time of the PTCA and you see that after the PTCA is when really a lot of the events started occurring.

Unfortunately, in this trial we did not continue abciximab for more than an hour after the intervention, and that is when a lot of the events were occurring. So, we

didn't really even cover the appropriate time period perhaps with medical therapy.

[Slide]

Now going on to abciximab as an appropriate active control for PCI trials, basically I want to argue that there is consistent and substantial benefit across diverse trials. There are different patient populations. There is a high risk population in EPIC. There is a broad intervention population in both EPILOG and EPISTENT. There are three acute MI trials, RAPPORT, ADMIRAL and the Munich trial. There are different heparin regimens; there are different devices. These trials were conducted over a period of 8 years. They had approximately 10,000 patients in them. They also have very consistent endpoint definitions.

[Slide]

So, the primary endpoint in these trials, or one of the pieces of the primary analysis has been 30-day analysis of death, MI and urgent intervention. For the large trials, EPIC, EPILOG and EPISTENT, you see the results here. Now, death definition is obvious. MI, we consistently required multiple measurements with at least a 3 times elevation of CKMB or, in the absence of MB, measurement of total CK. Urgent intervention generally would require recurrent ischemia requiring intervention, usually the criterion we would use was within 24 hours of the ischemic

event.

So, you see fairly consistently, in EPIC about 700 patients per arm, about a 4.5 percent reduction in death, MI and urgent intervention at 30 days. This is a high risk population. EPILOG, about 900 patients per arm -- here we had 2 heparin strategies that had very similar endpoint rates so I combined them here. There is over 5 percent reduction in death, MI and urgent intervention at 30 days. That is highly significant. In EPISTENT, again this is a broad population here, these were patients who were amenable either to stenting or to PTCA. In the placebo group all patients got stent. There were two abciximab arms, one received abciximab and PTCA; the other received abciximab and stent. The results for this endpoint were nearly identical and, again, you see close to a 5 percent absolute reduction in death, MI and urgent intervention. Obviously, the relative reduction is about 35 percent here and over 50 percent here in the EPILOG trial, and a little under 50 percent here, in the EPISTENT trial.

[Slide]

We have also conducted three trials in patients receiving direct angioplasty for acute myocardial infarction. In ADMIRAL the primary endpoint was death, MI and urgent revascularization. The secondary endpoint was studied also in the RAPPORT and the Munich trial, conducted

by Franz Joseph Neumann.

There were 500 patients in RAPPORT, death, MI and urgent intervention, reduced again by absolute 5 percent, and that did reach statistical significance, less than 0.05. In ADMIRAL just 300 patients, about an 8 percent advantage in death and MI, 300 patients that reached statistical significance, 0.02. That was the primary endpoint. Neumann, this trial really wasn't powered for a clinical endpoint but that endpoint, death or MI or urgent intervention, was reduced substantially, by over 7 percent in this trial.

[Slide]

There has been a lot of discussion of mortality here, and we would like to propose that there actually is some reasonable evidence for a mortality benefit with abciximab.

First of all, just to remind you how this works, basically the one line means that the placebo and abciximab would be equivalent for mortality and the vertical dash is the estimate of the hazard ratio for mortality with abciximab relative to placebo. Left of that one line indicates that there is lower mortality with abciximab.

What we did in this analysis, and I can discuss related analyses for those who don't like this particular one, is to look at all follow-up for patients who received the most commonly used IIb/IIIa inhibition regimen, which is

a bolus of abciximab immediately prior to intervention, followed by 12-hour infusion. So, we used all follow-up and we studied in this analysis patients who got the same intervention. So, balloon is being compared to balloon within each trial and stent to stent within each trial.

In each trial you got consistent results. The combined results suggest that the hazard ratio for mortality is actually 0.69, with a p value of 0.006. This is intent-to-treat analysis, and the results are for 3 years in EPIC, 1 year for EPILOG and EPISTENT. We had 6 months of follow-up for the remaining trials. Basically, the results were consistent whether balloon was used or stent was used.

[Slide]

In terms of when the mortality benefit accrues, this is a combined analysis of EPIC, EPILOG and EPISTENT, which are the trials where we have 1 year of follow-up. So there are 600 patients essentially in this analysis. You can see that there is a little bit higher mortality rate immediately, in the first couple of weeks, but most of the mortality actually accrues after the first couple of weeks although it is occurring at a slower rate.

You can also see that there is a slight divergence of the curves, say, in the first month but there is continued divergence of the curves after that. If you actually divide the analysis into the first 2 weeks versus

later, the hazard ratio is 0.67 in the first 2 weeks and 0.69 after the first 2 weeks, and the p value -- again, this is a subgroup analysis, after 2 weeks does go below a nominal 0.05 level. So, that may not be conclusive evidence but certainly it is very suggestive that there is some late benefit here with abciximab.

[Slide]

So, basically, we feel like we have shown consistent substantial benefit with abciximab in PCI trials. There is a 30-day benefit that is maintained through 6 months, 1 year, in EPIC 3 years. I haven't shown the other endpoints for that. And, there is a consistent long-term mortality benefit across the trials.

[Slide]

This shows the longer term follow-up that we are aware of in the small molecule trials that have been presented today -- mortality at 6 months. Basically, in 3 of the 3 trials the mortality was slightly higher at 6 months, but essentially it is identical. So, is there any real difference between the drugs that may affect long-term outcome is an interesting question to us.

[Slide]

The things that may cause a difference between these, we feel, would be the unique abciximab pharmacology, and that there are unique receptor binding characteristics

due to E3A but also to alpha-v-beta-3 and Mac 1. Abciximab inhibits not only platelet aggregation through IIb/IIIa but also thrombin generation with IIb/IIIa and alpha-v-beta-3 inhibition, possibly it inhibits atherogenesis and angiogenesis by alpha-v-beta-3 inhibition, and inhibits inflammation by Mac 1 inhibition. But, finally, there is unique gradual and tapered recovery of platelet function with abciximab in that abciximab is something whose half-life on the platelet is measured in days while the small molecules are measured in hours.

[Slide]

So, the rationale for the use of abciximab as an active control in PCI trials is the unique pharmacology, consistent substantial results across diverse trials at 30 days and long term, possible mortality benefit, and we feel that in PCI trials it is really not adequately demonstrated with the other compounds when patients are undergoing immediate PCI. So, we think it is reasonable to use abciximab as an active control in future PCI trials.

[Slide]

To actually give a practical recommendation in this regard, we combined three trials with a substantial amount of data to look at the death, MI, urgent intervention endpoint again. You see the same 5 percent-plus absolute benefit that basically you have seen consistently in all the

large trials and in the small trials. When you look at the hazard ratio, it is about a 50 percent reduction in the hazard ratio, and the confidence interval for that, the upper limit, is 0.61. So, the hazard ratio estimate is actually quite tight and the benefit is quite substantial. This is the ideal sort of situation in which to do an active-controlled trial.

[Slide]

Just in terms of a practical recommendation -- and these are assumptions that can be changed, obviously, with slightly different recommendations, abciximab as an active control in a PCI trial, the population would be immediate PCI. From our trials, it does not seem to matter which population you use in immediate PCI, you get the same 30-day results for death, MI and urgent intervention. Pretty consistent, you get around 11 percent event rate. Again, we did use uniform screening of enzymes. I think that would probably be important.

We would also propose that there should be a secondary endpoint of 1-year mortality to exclude possible mortality increase relative to abciximab when there has been enough evidence starting to accrue that there is a possible mortality benefit. It is very easy to go back at 1 year and measure mortality with a phone call.

In terms of sample size computation for this

trial, if you assume that the new therapy is equivalent and you want 80 percent power to show retention of 50 percent abciximab benefit and, again, if you want to show less retention you can get a smaller sample size, you would need 2800 patients per group. So, this seems to us like a fairly straightforward proposal for an active-controlled trial in a PCI indication.

[Slide]

In summary, as has been discussed all day and there doesn't seem to be any consensus about any way to do an active-controlled trial for ACS trials where PCI is not immediate essentially, we feel again that abciximab is the appropriate active control for PCI trials, again, due to possibly its unique pharmacology, consistent substantial 30-day benefit and the possible long-term mortality benefit that has been suggested. Thank you.

DR. PACKER: Thank you. It may be recognized by all who have been listening to the discussion this afternoon that each of the sponsors' presentations highlights a different issue as regards to the conduct of an active-controlled trial. Dr. Kitt's presentation focused on delineation of the appropriate patient population and the syndrome, and Dr. Sax's presentation focused on the issues related to definition of endpoints, and this particular presentation, Dr. Anderson's presentation, focused on the

selection of a comparator agent.

In that spirit, before we go over this issue in detail, let me ask Dr. Reid to present his meta-analysis which, in fact, deals with the same issue of the selection of a comparator agent, and we will bring both of these presentations up for discussion.

Meta-Analyses

DR. REID: Dr. Packer, ladies and gentlemen, thank you.

[Slide]

Since being invited here and realizing the nature of this discussion, we decided to change the name as shown on this slide from "Yet Another Meta-Analysis" to the one you see there. We thought it was a little more specific.

In this, what we will do, through mechanisms that I suggest would create discussion, is compare the various, in this case three, GP IIb/IIIa receptor inhibitors.

[Slide]

While I am presenting the data, I am deeply indebted to the team with whom I have worked, the internal team at Eli Lilly and the external part of our team at Metaworks. We were specifically using and working with Metaworks to provide independent outside statistical consultation so as to help validate those preliminary results which we felt applied to the conclusions which you

will see.

[Slide]

The purpose then of this talk is to compare and contrast the efficacy of parenteral GP IIb/IIIa inhibitors in the management of PCI patients. This will then be restricted to those agents which are used in patients clinically today, that is, those so-called FDA-approved parenteral agents.

[Slide]

Let me start with the presentation outline. First, these findings, we would suggest, will show consistent results using various statistical methods in PCI patients at 30 days with the endpoint of death and myocardial infarction. The second conclusion will demonstrate that these analyses will show that abciximab appears to differentiate itself from the other agents.

Now, our purpose is not to display a variety of powerful statistical techniques but more, speaking as a clinical trialist, I would offer to you that it impresses me when one can use a variety of statistical techniques that appear to be giving you the same result or trends in the same direction time after time. I would certainly leave it to the experts to prioritize which of those is more important.

[Slide]

Every one of us who had the opportunity to speak with you this afternoon have all talked about heterogeneity and we are no exception. Since you have seen so many of these slides presented, I probably will run through them a bit quickly, if I may.

The first point is that we can recognize that drug dosage in any trial, and particularly when comparing across trials, can create heterogeneity.

Secondly, the patient population studied, and this has been amply illustrated, talking particularly between ACS and PCI patients where, within that diagnostic category we may have additional obfuscation appearing in the form of disease definition -- regional differences. For example, you saw the data from Europe and the United States. Or, inclusion and exclusion criteria, particularly those that are what one may call more subjective.

The next point is that the selection and the timing of the endpoints, something that we have talked about and I think we are going to hear more about at the end of this next discussion -- these two things can influence the heterogeneity.

Finally, the differences among agents, the mechanism of action and, as Dr. Anderson has pointed out, abciximab, in contrast to the other two agents, appears to inhibit the GP IIb/IIIa receptor by a mechanism which is

referred to by our chemists as stearic hindrance, that is, it is not specifically binding to that receptor and, while all three affect the same result of platelet inhibition, it appears to be doing this by a different mechanism of action.

In addition, as Dr. Anderson showed you, it has binding to the Mac 1 and victronectin receptor which differentiates it clearly from the other two smaller molecules.

Finally, perhaps these mechanisms of actions and other effects appear to provide substantially differentiating pharmacodynamics when one compares abciximab with the other two small molecules.

The result, of course, of all this list could be two things, efficacy differences or inability to detect these if the heterogeneity is not controlled.

[Slide]

We, then, undertook, in conjunction with Metaworks in Boston Massachusetts, a meta-analysis. What I want to do is to try to summarize for you the methods that we used to collect these data. I would hasten to add that the data that we will be presenting to you must be considered preliminary since the analysis is currently under way. These are merely stopping point and then bringing out the data as they appear now.

First, we prespecified the experimental

hypothesis, which will be summarized for you. Secondly, we prespecified a study design, such as you would with any clinical trial. Thirdly, we pre-wrote a protocol and a case report form before any data were collected. Fourthly, we prespecified the statistical plan which was to be used in the analysis of the data once they were collected.

[Slide]

Fifthly, we prespecified the patient population, and when one adds this all up and looks in the ICH guidelines, you find that this meta-analytical plan will be consistent with the statistical principles of the ICH guidelines.

[Slide]

This, then is a summary of what we did in terms of the methodology, as well as some of the features that would then be under these various labels. The total sample size for this analysis was 13,350 distributed across the three agents, as shown on the first row of this slide. The patient population in order to reduce heterogeneity, from a statistical perspective, was limited to PCI. I would add that it also allowed us to exercise two other important features. As we look across the trials that we have reviewed today, all agents have demonstrated their best odds ratios in PCI trials. So we felt that this really gave everybody an equal opportunity to sort of show his best.

Thirdly, from a pathophysiologic perspective, and not unrelated to the first two, it would appear that one of the possibilities of benefit that is derived from IIb/IIIa antagonists is that they prevent the events that PCI induce. Thus, it allows one to derive benefit which otherwise may have been deprived if that agent were not present.

The dosage in these trials that we reviewed was consistent with current labeling. We chose an objective endpoint, as objective as we could, across these trials, which was death and myocardial infarction. We fixed it to a single time point for purposes of the interim analysis to 30 days, and the number of studies is shown at the bottom.

[Slide]

So, these now will be the results. If we look at what comes in when one talks about the effects with the entire class or group of agents, we have the first point and this is the so-called combined effect, shown here with its odds ratio of 0.6, and then the confidence intervals ranging from 0.49 to 0.73. Much as was shown in the previous presentation, we would have 1 then as showing no difference between placebo and the treatment. To the left of this line would be favoring therapy, or to the right actually would be favoring placebo.

So, from this part we can conclude that we have redemonstrated that which has been shown by previous

analyses, and that is, there, indeed, appears to be among these three treatment agents a significant effect in favor of the therapy that is employed.

The next point shows you, as we broke it out now for tirofiban, we have an odds ratio of 0.68; the ranges are shown here. The next agent is eptifibatide is shown here, with an odds ratio of 0.82; the ranges shown to the right of that. Then for abciximab we have an odds ratio of 0.46, with the confidence intervals as shown here.

The next step that we wished to take was to ask the question how would you compare these and what conclusions could you reach? In order to do this, we undertook the analytical technique which will be shown on the next slide.

[Slide]

First, we performed an ANOVA to do paired comparisons, and when we did that we found that abciximab against tirofiban gave us an ANOVA of 0.02. Next, we compared it against eptifibatide and found an ANOVA value with a p of 0.001. Finally, tirofiban against eptifibatide gave a p value of 0.156.

[Slide]

Because of the statisticians' concern of variability among the controls, we then undertook one additional analysis, an analysis of covariance, and

repeating these same types of comparisons between abciximab and tirofiban we derived a p value of 0.048. When we compared abciximab against eptifibatide we obtained a value of p equal to 0.002, and finally tirofiban against eptifibatide gave a p value of 0.082.

So with these two types of analyses of a subset of the data, it appeared to be consistent among the p values which suggested that what was shown in the previous slide may, indeed, point to an effect that differentiates abciximab within the combined group effect.

[Slide]

Finally, being a clinician I always like carry-away messages -- how would this translate to how many patients do I have to treat to prevent something I don't want them to get? So, fortunately, our statisticians came up with what is referred to as NNT or, simply expressed, it is the number needed to treat to prevent one event. In this case, the endpoint is death or myocardial infarction at 30 days.

[Slide]

When we do this -- shown on the ordinate is the NNT or the numbers needed to treat to prevent an endpoint, we find that one must treat 23 patients to prevent death or myocardial infarction in 1 patient or, with tirofiban 38 or, with eptifibatide 67 patients, these again being compared

against placebo.

[Slide]

In conclusion then, first we suggest that the combined group effect of GP IIb/IIIa inhibitors shows a distinct decrease in death and myocardial infarction at 30 days compared to placebo therapy in PCI patients.

Secondly, abciximab appears to show a significantly greater decrease in death or myocardial infarction at 30 days when compared to either eptifibatide or tirofiban. Thank you for your attention.

DR. PACKER: We are going to open up committee discussion on both Dr. Anderson's presentation and Dr. Reid's presentation at the same time. Could I ask both David Kong and Tom Fleming to address the methodology used in the meta-analyses, the issues that they would like to raise in terms of the techniques used or the conclusions reached? David, do you want to start?

DR. KONG: Yes, I think that I have already sort of described my position on what to make out of indirect comparisons. I think once you start shrinking the available number of patients that you have in each of the groups to what is represented by the individual compounds, the variability in the data will enlarge in order to make indirect comparisons among agents.

Certainly, I would commend the use of a random-

effects model for this type of analysis. A random-effects model assumes that each of the effects falls along some certain distribution. So, in terms of attempting to incorporate the heterogeneity and variability among trials, I think that is certainly appropriate. However, yes, I still do have a deep concern among indirect comparisons amongst agents.

DR. PACKER: Tom?

DR. FLEMING: Certainly, it is complex as we try to glean as much as we can from current data, and there is a strong interest in wanting to be able to compare agents in A versus placebo and B versus placebo assessments to be able to say something about A versus B. It is intrinsically difficult and, obviously, the larger the signal then the more confidence one has that there is a difference.

The question that I might ask the committee to consider as they are thinking about this is to what extent were these trials the same? If we are pooling the data predominantly over four to five different studies, are these studies really comparable in terms of their patient populations, in terms of the manner in which the endpoints were defined and assessed and monitored? Were the quality of the data in the trials consistent? These are among the issues that need to be considered and understood with some considerable confidence in order to be able to justify a

conclusion that A is better to B when A was compared to placebo and B was compared to placebo.

DR. PACKER: Let me just make sure that we can define why we are having this discussion. We are not here to provide any sponsor with the opportunity to claim that their agent is better than any other. We are here entirely to address the issue as to whether any comparator agent for an active-controlled trial can be selected with some degree of confidence.

If I understand it correctly, Dr. Anderson's presentation primarily made the point that you would suggest that the consistency of the data was greater with abciximab than with others, therefore making it your preference. Dr. Reid's presentation was not so much based on consistency but based on superiority as opposed to consistency. Is that a fair representation and summary?

DR. REID: I think that is fair, Dr. Packer. I would just add one other feature, and that is one cannot simply reach in the GP IIb/IIIa basket, pull any agent out and expect to get the same efficacy.

DR. PACKER: Okay. Let us go through the committee concerns. Dr. Kitt, I promise you, you will have more than ample opportunity but I want to get the committee concerns on the table first. We will begin with Ray, and we will go to Marv, Udho and Rob.

DR. LIPICKY: Well, I guess I would just like to say what I usually say, and that is, you know, if we ignore all of the stuff about meta-analyses and are the populations the same, and everything else, as a single trial to sort of come to the conclusion that something really has been shown you have to think about p's at 0.00125. I didn't see anything -- sort of close maybe. My understanding of when you start believing meta-analyses is when another zero gets added before the significant digit, and then you might start paying attention.

So, fundamentally, I haven't seen anything that would make me think that there was a difference between agents, nor that there is nothing different between agents, and I will leave it at that.

DR. KONSTAM: Let me just say, I mean, I think that Dr. Anderson's and Dr. Reid's presentations are powerful and make a very good point that abciximab is an adequate agent for being employed as an active control, and I accept that. I accepted it before their presentation and I am, if anything, strengthened by their presentation.

Getting beyond that I think is where we get into trouble. I was convinced earlier by Tom Fleming that comparisons across these agents at this point in time, or maybe at any point in time, in the absence of head-to-head comparator data, is treacherous. I don't know whether Tom is

getting tired, or what it is, but he didn't seem to come across as clearly about it in his statement a moment ago, but I think it is extremely treacherous across these trials. I think these trials are enormously different from trial to trial, population to population, dose regimen used, endpoints used. So, from a general perspective, I don't get much of anything from comparisons across them.

But I just want to take that a step further and to an extent just comment on the direction that things appear to be in relative to them, I would propose, and ask for comments to refute it, that the findings could represent more effective, a more aggressive antiplatelet regimen. I think they pointed out a number of points about abciximab, one of which is that it has a very long half-life, and I think platelet aggregation in these settings is bad.

Now, if that were the source of some great effect, if we believe them, I think one would also see high rate of bleeds and, in point of fact, what I didn't see in your meta-analyses is a comparison of the relative major bleeds across the different groups of agents. I am sure you are going to pull out slide number 432 -- actually, I would like to see it. But, to my reading of the literature, having not done a meta-analysis or comparison, I see a significant increase in the rate of major bleeds in the EPIC trial and I don't see it clearly in the other studies as well. So that

would be what I would propose as a possible explanation if I were to believe the differences.

DR. ANDERSON: We do have a slide but I don't know the number of it.

DR. PACKER: Remember, the issue we do not want to discuss is whether the agents are different, materially different either in efficacy or in safety. The issue that we want discuss is whether the data that exist now allows us to identify a comparator agent with confidence.

DR. KONSTAM: I am comfortable with that. I think, to be fair, Dr. Anderson's and Dr. Reid's presentations went beyond that. So, we could either just accept that we are not going to talk about it or say that we are not sure that we believe it.

DR. ANDERSON: Just very briefly in response to your point, we have found that really bleeding is not necessarily associated with how effective something is and, in fact, where we have had the best efficacy results, in EPILOG and EPISTENT, we had the lowest the bleeding rates and we had lower than placebo in EPISTENT and I believe the low-dose heparin regimen. So, if anything, we would claim exactly the opposite.

DR. THADANI: You showed good data. I don't think you can compare the agents because some patient populations are different in some of those studies. Some have included

patients with a recent MI and some studies not.

What I am struck by, which you did not conclude, is that as soon as you blow the balloon up or put a stent in you are driving your enzyme-driven infarct rate by at least 6-10 percent. So, what you are telling me is, okay, if you need a PCI and order the primary therapy you can prevent an iatrogenic infarct. Is my conclusion right? Because each of the three you showed, at point zero, in the placebo group it goes almost up to 10 percent. So, here you are telling me that I tell a patient, okay, if I am going to blow a balloon up I am going to cause an infarct, and I will give you a drug which is going to drive you down from 10 percent to 6 percent. I realize the benefit continues up to 30 days, and that is why I am having a problem when I ask do you accept MI as an endpoint just driven by enzymes. That is the difficulty I have as opposed to the natural history of a disease, because if you translate this that every patient with ACS has to go intervention like this, you are producing a lot of infarcts which you are treating the patient to prevent it.

DR. ANDERSON: No, I don't think we are suggesting they should all go to intervention, but that --

DR. THADANI: So, what you are saying is if you have to go for intervention, this is the way to go? Am I right?

DR. ANDERSON: I am saying if you go to intervention and, if at that point, you are making a decision to give a IIb/IIIa inhibitor, abciximab is very useful in that setting as a potential active control.

DR. THADANI: But not for ACS alone?

DR. ANDERSON: For medical therapy, you know, we didn't feel like there is an appropriate active control at this point.

DR. SEIGEL: I want to extend a little bit the observation or the comment you made a couple of times, Dr. Packer. It is not on the table whether one of these agents is superior. It is on the table which could be used as a positive control. It may be on the table, and this has been unclear to me from the wording of the papers and from the discussion, but one thing that may be on the table that is addressed here is could you use a class-specific group estimate of effect size to estimate the effect size for the purpose of planning a clinical trial in which one member of the group could be used as the active control, something that has been done, for example, in thrombolytics?

There, I would say that if you believe that one can make even a plausible case -- not proof, but if you can believe that one can make a plausible case that there are real differences in effect size, then you have to ask very seriously whether you would want to get a pooled estimate of

effect size from different therapies and then apply that effect size to an assumption of any one therapy within the class.

DR. PACKER: That is actually a question that we should focus on for a moment, but before we do, Rob?

DR. CALIFF: I think Jay has already gotten a large part of the way that I wanted to go, which is to get away from comparing the drugs to see which one is better, and more try to generalize the issue which we will see over and over now, that you are developing a new therapy and you are in a field where there is a bunch of therapies out there that are already being used. So, how do you approach it?

And, the only part of the presentation which I thought I really disagreed with was the use of the term "prespecification" because I don't think the ICH says that you do the trial without a hypothesis, look at the results and then specify what your hypothesis is. And, there is no way to do a meta-analysis of trials that are already completed without knowing what the results are ahead of time sort of qualitatively. So, it is not really prespecified the way we talk about an experiment, unless you prespecify before you do the individual trials and have a plan for how you are going to combine them before you have seen what the results of each individual trial are.

So, the generalized issue there is when thinking

of your strategy in a positive control trial, like everything else we have said, it is a matter of sort of taste -- which studies do you include or not include? We have been through this we ACE inhibitors, for example, with IRBs. There are a lot of different ACE inhibitor versus placebo trials in different types of heart failure. You can include or exclude various trials. And, if you already know the results you might have a tendency to prespecify the trials that you really liked.

The second point is that there is a problem here in terms of selecting one of the agents that is not intuitively obvious, I don't think, and that is if there is heterogeneity in your analysis, which this analysis clearly shows and I think it is very well done, if you select the one that shows the greatest effect it may seem that that would be the most difficult obstacle in terms of a non-inferiority trial. But when you get into the putative placebo argument, in fact, if you select the one that has the least effect it is harder to show that you are actually different than the putative placebo. So, it would see like when you first look at it, you will take the most difficult choice and that is the hardest thing to do but, in fact, if you take the least effective agent in your meta-analysis that is the hardest one to show you are better than placebo, and you may have a higher chance to show that you trend in

the right direction compared to the active control. But since the confidence interval butts right up against no effect in the least effective agent, you have a difficult time in your putative placebo argument, which we haven't gotten into here but becomes a critical part.

Lastly, I don't think biological differences should be any part of the discussion today because show me a result and I can give you a biological difference that might explain it. I think when it comes to this kind of an analysis we have to be talking about the outcome data.

DR. PACKER: Let me see if we can move the discussion forward the way that we want to. Tom, I know you probably want to say something in your own right but could I ask you to deal specifically with the question as to whether an appropriate comparator here for any active-controlled trial should be either one agent or a pooled estimate. The one agent issue Rob has already outlined has certain advantages or disadvantages depending on which agent you use. The other question which Jay brought up was is it more appropriate to use a pooled estimate. If this committee three years from now were to see a trial in which a comparison was done using a pooled estimate or a comparison was done with a single agent, what are the considerations we should be worrying about three years from now that would influence the design now?

DR. FLEMING: That is what, in essence, I wanted to address. I would though like to just briefly endorse again Rob's comment about prespecification. It bothers me greatly when we have all the studies in hand, plan a meta-analysis -- we know what the results are and, obviously, you want to do as best you can to prespecify but it is not the same as having specified a hypothesis before any of the data were in hand.

The issue that you raise, Milt, I would like to comment on. We have made the point that if you have an array of studies that look at A versus placebo, B versus placebo that one has to be extraordinarily cautious in using those data to conclude relative efficacy of A versus B.

Having said that, that doesn't mean that in the absence of having considerable convincing proof that A and B are different that the net fall-back is that A and B are the same. You have just as much difficulty in proving that A and B are the same. So, if I have an array of studies that look at A versus placebo, B versus placebo, and I have decided to choose arm B for my active control, the fall-back isn't to presume that I can estimate the efficacy of arm B with the global analysis because the burden of proof is on me when I do that to be able to conclude why the efficacy of A versus placebo is reliably giving me further insight about B versus placebo. If I believed, if I truly believed that it was

highly likely that A was better than B and I was going to follow the paradigm that is evolving here for an active control study, I would want to pool the information from A and B, get an inflated estimate of efficacy, then I have to preserve half of that level of efficacy, then I want to compare myself against B which is much easier to beat than A -- if I truly believe from looking at the data that there was a difference in efficacy between A and B.

So, it really is important here -- the fault here is if I am going to use B as my active comparator, B data is what I have to use from past experience against placebo to estimate B. The burden of proof is on me to establish why A data is as relevant as B data when I am trying to estimate efficacy of B.

DR. LIPICKY: But there is another set of considerations that you might want to comment on. For example -- and I may be wrong in the assertions that I am making now -- if you have concluded that there is ample evidence that there is a class effect and you have not been able to conclude that there is a difference between the members -- you certainly ca n always say, well, there could be but you can't conclude that there is, I would make the argument that your best estimate of the population's treatment effect is, in fact, the meta-analysis of all of the trials versus placebo.

DR. FLEMING: Of the class.

DR. LIPICKY: Of the class, and that in fact the confidence limits beyond around that point estimate are as small as they can get. That is your best estimate of the treatment effect of the class. So, I don't see that it is necessary to choose A or B. Your control could be all of the members against new drug because you really, in fact, have your best estimate of the class effect and that gives you your best confidence limits and your best position. So, that would just mean that your control arm would include randomly all of the members of the class.

DR. FLEMING: I have no concern if you done studies and you have established a class effect, and those studies have been incapable of definitively concluding that A is better than B in that class. I have no problem if you choose A as your active control, B as your active control or a combination thereof.

My concern is by virtue of your inability to prove a drug difference if efficacy within that class, that doesn't lead to the conclusion that they are the same. You may be under-powered. There may be true differences that don't reach your 0.00-whatever difference that you are suggesting you would need to see. And, there are relative degrees of confidence that there may be heterogeneity. If one looks at the data and believes that it is entirely

plausible, if not likely, that within this broad class effect there is drug-specific effect as well, in those settings I am uncomfortable attributing the entire class effect to one of the given agents.

DR. LIPICKY: Right, but in fact uncertainty of then saying I am going to choose an agent -- there is an alternative to choosing an agent from the class, and the uncertainty in choosing an agent would be you have a trial or maybe two trials that, in fact, estimate the magnitude of treatment effect and the variance that is associated with it. So, you have less certainty there. You may have more certainty -- you may be more comfortable because you are not sure there is a difference between drugs, but if you choose a drug then you have more uncertainty in this vector that you are chasing for establishing the non-inferiority. So, it is a trade-off and it is not clear to me where that trade-off sits.

DR. FLEMING: I agree. I guess the bottom line of what I would want to come through with here is lack of evidence of a difference is not evidence of a lack of difference.

DR. LIPICKY: Right, but there is a question that can be addressed and should be answered with a yes or no, that is, does what we have looked at today fit in the category of no difference has been shown so there is

reasonable comfort with saying that it would be okay to choose all of the controls, or is it the statement that you are making, that no difference doesn't establish that there is no difference?

DR. FLEMING: My summary is that there is strong evidence for a global effect of the class, and we are far, far short of having adequate data to conclude that all of the agents within the class are equally effective.

DR. PACKER: Let me see if I can understand the implications of the distinction. Ray, what you are proposing is that if one is going to use a pooled estimate one could, as a control group, use all three agents in some randomized fashion in a manner similar to the cooperative nitroglycerin study.

DR. LIPICKY: Right.

DR. PACKER: On the other hand, if a sponsor thought that that was too complicated the alternative was to select one agent, and Tom's argument is if you are going to do that you shouldn't use the pooled the estimate, you should use that agent's estimate to do that.

DR. LIPICKY: Right.

DR. PACKER: Is there any disagreement between the two of you on this conclusion?

DR. LIPICKY: No, absolutely none.

DR. PACKER: Terrific. Please identify yourself.

DR. OLKIN: Ingram Olkin. I am from Stanford University. I think the points raised are actually pretty treacherous in terms of making conclusions but, to put it in the vernacular, if the Yankees are better than the Braves and Cleveland is better than the Braves, it is very difficult to compare the two.

But I would like to focus on what Tom's point. Statisticians have developed a technique called analysis of variance, and what the analysis of variance does is it does give you an overall effect, an overall mean effect, and then it gives you an effect due to each of the agents. Now, that was not done in the Kong et al. study, namely, the designation of the agents was not really taken into account. It was a meta-analysis of all. And, that is actually fine. People do that and it is a legitimate point. However, the analysis of variance does both. It is a more powerful technique than just combining all because it does give you the overall effect regardless of the three agents. Then it tells you whether there is a difference between agents. For that, if there is a difference between agents, you can do multiple comparisons which will make comparisons. The key problem is the comparisons are not between an agent and an agent but each agent versus its placebo. That is the rub.

So, in order to resolve that, you either have to have some hypotheses; you have to build a model, or you have

to do something in which there is a direct comparison. So, I think the point is I am not disagreeing, and I think the statisticians would agree that analysis of variance is probably a better procedure.

DR. FLEMING: If we were in an ideal world, we would look at three experimental agents in the same trial against a common placebo. That way, you only have to put one-fourth of your population on placebo and you get placebo comparisons of each and you also get direct head-to-head comparisons where you have more standardization of populations, intervention, definitions, etc. That is not the world that we are in with the data that we have, as I think Ingram is getting at, and that is the essence of where we have a much weaker basis to conclude that each of these agents is providing essentially comparable effect to the global class effect.

DR. PACKER: I am going to ask Jeff to address his question. I want to take a poll of the committee and then I want to give Drs. Kitt and Sax the last word.

DR. BORER: There are two issues that I would like to raise. First of all, we are talking now about pooling, or possibly pooling data, or not pooling data for a comparison. I think it is important not to lose sight of the point that Marvin raised earlier and that I certainly agree with, which is that we really are looking at two very disparate

conditions. So, if we are going to talk about comparisons, pooled or unpooled, whatever, we really need to be thinking about PCI and ACS separately.

The other point is that although intuitively, barring some reason not to do it, it seems appropriate to think about or move towards pooling all the data to have a more stable point estimate, if one does. We are talking here about three agents that have, it seems to me, very importantly different molecular structures and pharmacological effects. That is, when you take eptifibatide and tirofiban on one side and abciximab on the other side, there seem to be important differences here. So, I wonder if it is reasonable to call these agents a class altogether. It seems that that is a potentially important confound. Forget about the fact that the data look the same, or they don't look the same, or there seems to be a difference but that doesn't mean there is one or there isn't a difference. The molecular structure and pharmacological effects seem to be importantly different. I would be concerned about pooling at this point given that fact.

DR. PACKER: But you can never underestimate the ability of sponsors to describe differences. We have seen that with beta blockers and ACE inhibitors.

DR. BORER: You can describe differences in pharmacological effects just by looking, but when you talk

about a small peptide and a big antibody, to me intuitively it seems like you are talking about two different species.

DR. PACKER: Let me get a sense of the committee. Would anyone -- and this includes our invited guests -- would anyone object if a sponsor wanted to design a trial that would compare their drug to abciximab assuming that all the other issues of a positive control trial could be addressed? Would anyone object to that comparison?

DR. GRINES: I still think there is a major role for placebo-controlled trials. We have discussed this a little bit before but there are a lot of trials you can design to make it look equivalent by patient selection or concomitant medications.

DR. LIPICKY: That is correct. No one has said you can't do a placebo-controlled trial. But the exercise today is to try to figure out how you do a positive controlled trial.

DR. GRINES: I wouldn't object to a positive controlled trial as long as you had a placebo-controlled trial to show efficacy as well.

DR. PACKER: I understand.

DR. LIPICKY: No, no, no. That is not the name of the game today. It is a positive controlled trial from which you would conclude that this drug works. How would you do that? And, the question that Milton asked was if you did a

positive controlled trial that, in fact, would be of approval quality and you chose abciximab who would object? That was the question.

DR. PACKER: Right. Cindy, I have to help you because I am going to assume that all of the issues you are worried about, that have nothing to do with the selection of the drug, will be adequately resolved to your satisfaction. That may be impossible.

[Laughter]

But there is another question that is coming right after this. If one could resolve all of the levels of uncertainty that you have, would you object to a positive comparison trial with abciximab, using the abciximab point estimate and confidence intervals? Anyone who would object to that? No one would object.

Would anyone object to a similar trial being done with tirofiban, using the tirofiban point estimate?

DR. THADANI: I think there are problems on the whole because if you are saying, as Tom said, you can't have one agent to another and if most of the data is driven by drug A, I think there are problems and I would like to see one placebo-controlled trial as well.

DR. PACKER: That is not the question.

DR. THADANI: I realize that. I am not going to buy that you can do it with any drug given because I think

most of --

DR. PACKER: Then abstain. The third question -- no one objected to tirofiban. Right? Oh, Jeff -- maybe someone did.

DR. BORER: My recollection of the tirofiban data was that they were importantly split between PCI and ACS data, and that leaves me with the concern that there may not be enough data in either one of those pools to support a reasonable comparison, whereas the abciximab seemed to be plunked in one area and that leaves me with a feeling, a believing that there is a more reasonable point estimate for a single entity disease.

DR. PACKER: That would support Dr. Anderson's contention, not based on differences on effect but his contention that the point estimate shows a consistency with abciximab which lends itself to solving the issues of a positive controlled trial more readily. Is that the point you are making?

DR. BORER: Right, for one particular disease entity.

DR. PACKER: Okay.

DR. CALIFF: Milton, you asked the question hypothetically and I wouldn't object, but what a sponsor would find would be that they couldn't do it because for PCI the confidence intervals for tirofiban actually overlap 1.

So, by the rules of the game, you couldn't create a case that you could hypothetically beat a putative placebo.

DR. PACKER: Fine. Then let's turn to the next question which is the point estimate for eptifibatide doesn't overlap 1 --

DR. CALIFF: But it would take a large number of patients.

DR. PACKER: But it would take a very large number of patients.

DR. THADANI: Are you talking about an ACS trial or are you talking about --

DR. PACKER: No, PCI.

DR. THADANI: The PCI trial, if I remember correctly, the high dose didn't work, the low dose worked, which we haven't seen today.

DR. PACKER: It is irrelevant to the discussion.

DR. THADANI: I realize that. But if you are driving the PCI trial from the placebo data, it is very different patient populations from the primary PCI trials. Primary PCI trials I don't think have the confidence to show -- the doses used in PURSUIT are totally different than were used in the PCI trials. So, I don't think you can even discuss that, or I would have objection to say that you have enough data on the PCI group. If you are going to use the PURSUIT data -- okay, I buy that you can't do a mortality

trial, and if you can define how you are going to define infarction, then you have to have a huge sample size because even PURSUIT took -- what? -- 10,000 patients.

DR. CALIFF: So, Milton, there are really only two practical options. I mean, Keaven's plan I thought was quite rational in saying the other option is to pool the three --

DR. PACKER: And go against all three.

DR. CALIFF: Yes.

DR. PACKER: According to Tom's suggestion that if you are going to use a pooled estimate you need to against all three agents as the comparator. This is the result of Tom's and Ray's discussion. Jay?

DR. SEIGEL: I guess implicit in the comments about different doses, or ACS and PCI is the notion, and I assume it is implied in your question and the answers, that the conditions from which the estimates of treatment effect are made should be, at least vis-a-vis importance, the same as the conditions of the active-controlled trial. So you should have similar drugs. You should have similar entry criteria. There should be some level of similarity in the way, the amount and timing of the introduction of important procedures, although if we are talking about PCI that is implicit, and those other factors that are generally considered to be important. With that combination it is a little bit hard to do because you would have to weight the

combination comparable to the weight of the data from different agents, and then you would have to manage the patients in a way that reflects a lot of different management. But, conceivably, one could work through that.

DR. FENISCHEL: I just wanted to respond by saying what Rob said, which may be a little bit misleading -- it really is not true that an active control has to be distinctly better than placebo to be usable. All it has to have is some sort of defined position on the continuum. I mean, one could, for example, use as an active control a drug which was surely worse than placebo. It is just that one has to beat it rather more definitively than one has to beat placebo to show that one is successful.

I think one of the traps, one of the many traps of the unfortunate idea of preserving 50 percent of the benefit and so on, you know, when you try to look at that a little more closely you come up with ideas that some active control results are simply non-informative and, of course, that is not true. The example that Rob gave of tirofiban in PCI, where the confidence limit overlaps 1 -- suppose that that is true, well, all that means is it is still true that the point estimate is better than 1. It is still true that beating it is better than beating placebo. It is not a whole lot better than beating placebo but it is better than beating placebo. You go through the statistics and you are

better off than going against placebo as far as how good you have to be.

As I said, there are multiple pernicious effects of the formulation, for which I was partially responsible, of talking about preserving 50 percent of the benefit, and so on. It is a statistical dead end that leads into a variety of anomalies and paradoxes.

DR. PACKER: I think we have reached equipoise on the committee, but we do want to give the other sponsors the last word. We, by the way, never-ever give the sponsors the last word so this is unprecedented. Dr. Kitt?

DR. KITT: I will start by saying that it appears as though the subtleties of PCI versus ACS studies have again not been clearly understood, or at least have not been fully accepted or discussed during this presentation.

The meta-analysis that was presented -- there were assumptions in there, and there is always the problem of not having seen this information in advance but just reading from the slide, the meta-analysis took as assumptions that these were all PCI studies. Well, I don't know the actual details but I know how to do my math, and if I look at the Integrilin data, there were 5238 patients. If I subtract 4010, which is the entire IMPACT II study, I get left with the number of patients who were in the PURSUIT study.

Well, we spent the whole morning discussing that

PURSUIT was not a PCI study, and combining those results into these results doesn't seem to make a lot of sense.

The second assumption that was made in that meta-analysis was that these were package insert doses, but it looks to me that all of the IMPACT II information was folded into the data in that meta-analysis.

Along the same lines, conspicuous by its absence, I believe although I am not 100 percent sure is the CAPTURE study, but I will come back to the CAPTURE study in a minute when we talk about the mortality presentation.

I am assuming the same is true for tirofiban, by the way, but I don't know my numbers with tirofiban nearly as well as I know for Integrilin. But I want to just point out that this discussion that has just taken place is really about PCI studies and should not be confounded with studies in acute coronary syndromes. I agree with the original point that we have not really come to a conclusion as to what the control event rate or what the actual effect is with these studies because of the different designs.

I want to then turn to the mortality presentation that Keaven Anderson gave, where he compared the mortality benefit in all the PCI studies and then put up three studies, two with Agristat and one with Integrilin. Two of those three studies were not in PCI; they were in acute coronary syndromes. So, taking that information in that way

is sort of misleading.

Last but not least, I want to come to the CAPTURE study which was very conspicuously absent. I understand why it would not be included in there because the sponsor believes the dosing and the dosing regimen may not have been ideal, but neither was it for the other studies that you have already included. Just to be complete, I have the CAPTURE paper here. The 6-month placebo event rate was 2.2 percent. This is mortality. Whereas, with Reapro it was 2.8 percent. So, there was an increase in mortality at 6 months.

So, I just want to be sure that we have somewhat of a level playing field here. Again, my purpose is not to say one drug is better than another, but just to be sure that the information that is presented is equitable.

DR. PACKER: Dr. Sax?

DR. SAX: I don't want to comment on the analyses, except to say that I don't understand the mathematics either because there were two trials presented for PCI and the total N for that was half the size of the RESTORE trial. So, I think there are some methodologic issues.

Taking this aside, I guess I have the second to the last comment because the last comment always goes to the chair, just to say that I think there seems to be a consensus that it may be possible to an active controlled trial in the setting of PCI. But the issues are with

unstable angina, the acute coronary syndromes, non-St-segment elevation remain difficult, and I think that to the extent that there are difficulties it is going to require future sponsors to really look carefully at the things we have discussed -- the selection, the endpoints, in fact, interestingly, the make-up and design of the critical events committee which came out in the discussion today, and pull that together. I think those issues probably will have to be discussed on a trial to trial basis.

DR. PACKER: I promised the sponsors the last word and I will go in that direction. Can we move toward Dr. Throckmorton's presentation?

Timing of Endpoint Analyses

DR. THROCKMORTON: Thank you.

[Slide]

The topic of my talk is the timing of endpoint analyses, and if the previous speakers today have plunged fully into the heterogeneous, complicated, subtle and, my all-time favorite, treacherous sea of issues in designing active-controlled trials in IIb/IIIa inhibitors, I would propose to put my toe in gently and hope to avoid the sharks.

As you have heard, the trials used to support the approval of the three available IIb/IIIa inhibitors have employed primary endpoints ranging between 48 hours and 30

days and, in general, have shown beneficial effects on combined endpoints, including cardiac morbidity and mortality at the earliest time points measured, usually 48 hours, persisting with some variability out to 30 days. Some trials have, additionally, reported persistence for what I will call durability of efficacy through longer time points.

In addition, the meta-analyses presented have suggested, again, some variability in the results of the efficacy of these products between 48 hours, 30 days and perhaps later.

The timing of the primary endpoint has important implications for the size and design of any future IIb/IIIa inhibitor trial. Planning for possible active-controlled trials will require that we integrate the existing trials with their primary endpoints that vary and perhaps with varying durability of efficacy into a single effect size, with an ability to interpret this effect through time. A method for comparing this imputed control effect with a new drug effect at both the early and late time points, then, would seem to be desirable.

In my talk I will summarize four general patterns of data collection and interpretation that could be used for trials or have been used for trials of IIb/IIIa inhibitors. Then I will use data from some of the completed trials to illustrate a method of data interpretation that uses data

drawn from early and late time points to derive information not only about the acute effects of IIb/IIIa inhibitors in acute coronary syndrome, but to describe whether the acute effect is "durable" to later time points, in particular to 30 days.

I should emphasize that while this talk will draw on examples from the databases of the three approved IIb/IIIa inhibitors, this is not intended to compare between them or to reopen a discussion of their approval. The methods proposed have been applied in a post hoc manner in order to investigate the adequacy of this method to assess the durability of future IIb/IIIa inhibitors. This is intended to explore the consequences of applying this method to the available clinical database in order to aid the advisory committee in answering the questions posed to them by the division.

[Slide]

The first method would use a 30-day primary endpoint, a time well after the onset of the acute coronary syndrome. This approach has the advantage that demonstrating significant superiority at 30 days eliminates most of the concerns about the persistence of any short-term efficacy. In addition, statistical analyses methods in order to evaluate such are trial are in place.

The disadvantages of choosing this endpoint come

from the difficulties of showing significance at 30 days. Here, where the majority of the clinical effects appear to be in the first few days following administration of IIb/IIIa inhibitors, indeed in the first few hours after administration of IIb/IIIa inhibitors, the intervening time period serves only to add additional events to both the control and treatment groups, making demonstration of superiority more difficult and increasing sample sizes.

I should add, on the other hand, that in the context of an equivalence or non-inferiority trial the use of a 30-day endpoint might not be a conservative approach as differences between the two treatment groups may be obscured by the events occurring between the acute administration and 30 days.

[Slide]

The second type of trial would be to collect data only through 48 hours only, as has been discussed earlier today. This design has not been used for any drug that has been currently approved. Its advantages would be smaller sample sizes and, again, the use of standard analytical methods to assess superiority versus placebo.

The drawbacks to this approach are the smaller number of events that can be expected to occur and the lack of information about whether the treatment effect persists at longer time points.

[Slide]

A third method that has been used with some modification in two of the product developments would utilize an early primary endpoint between 48 hours and 7 days, and to examine the 30-day data for evidence that the difference between two treatment groups has not narrowed too much of an extent.

In this approach, however, no formal mechanism for determining if the 30-day difference is still clinically significant is in place.

[Slide]

Finally, and the method that I am going to discuss today calls for the demonstration of clinical efficacy at the earliest time point, 48 hours, followed by an analysis of the data at 30 days to make inferences about the durability of the clinical effect.

What I am going to discuss is the analysis of the endpoints in terms of a classic superiority trial design compared to placebo. However, this same discussion is relevant to an active-controlled design, equivalence or non-inferiority trial.

I should also say that there are other methods of assigning primary endpoints and collecting data apart from those that I have listed, and one of those is included in the questions for you today. My intent is to give an

overview of the types of the possible approaches to provide a context for the method that I will describe next.

[Slide]

The method in general is depicted schematically here. In this particular trial one can look at the difference in the event rate at 2 days, shown in the green, and at 30 days, and it is relatively apparent that there is no difference. That is, the difference in the treatment group, shown in white, and the control group, shown in yellow, is the same at 30 days and at 2 days. And, no one in this auditorium would have difficulties, I believe, in saying that the effect that occurred by 2 days has persisted through 30 days.

[Slide]

A greater difficulty is shown in this schematic, where the effect is clear at 2 days. That is, the treatment group has a much lower event rate, but this event narrows through 30 days, and the 30-day event rate difference is shown in green. In such a case, we have in the past looked at the shapes of the curves to draw inferences about clinical durability, and the method that I am proposing is a more mathematical approach, if you will, to this.

In broad terms, what I am going to propose is using the difference from the 30-day endpoints to derive an imputed treatment group from the control group at 48 hours,

from the early time point. If this imputed treatment group is then different from the control group at 48 hours, the interpretation would be that durability of efficacy has been suggested. I will go through an example next.

[Slide]

This data comes from PURSUIT and, as you can see, at 48 hours, the early time point, there was a significant difference between the control and the treated groups, with an incidence of the endpoint of 7.6 percent in the control versus 5.9 percent in the treated. At 30 days there was a 15.7 percent incidence rate in the control versus 14.2 percent in the treated, which achieved a p value of 0.043.

The method that I am proposing will take the difference in the event rates at 30 days, that is, 15.7 minus 14.2, and subtract it from the 48-hour time point to derive a 48-hour treatment group that will coincide with the difference in the event rates at 30 days.

Here is the math for that, 7.6 percent which is the event rate in the control group at 48 hours, subtracted from the difference in the event rates at 30 days, yielding an event rate of 6.1 percent. Multiplying that figure by the number of patients in the treatment group, 4722, gives you the number of patients in the imputed treatment group at the 48 hours that would have had an event. Again, the number is 6.1 percent. If you then apply Fisher's Exact Test to the

baseline control and the adjusted treatment group, one obtains a so-called durability p value of 0.004.

[Slide]

Schematically, this looks like this. At the early time point then, the first step, the trial has demonstrated clinical efficacy as suggested by the p value of 0.001. The 30-day p value is larger, however, the durability p value, computed as I went through, 0.004, suggests that there was clinical durability of efficacy through the 30-day time point.

[Slide]

There are at least two other patterns of results that exist in the database that we currently have that are worth going into briefly. First, from the PRISM-PLUS trial, if you look at the 7-day data and the 30-day data the percent reduction at 7 days was 5 percent, and this was nominally statistical significant. At 30 days the 3.8 percent reduction did not achieve nominal significance, greater than 0.05.

When the durability p value was calculated, however -- and, again, that would be calculated by subtracting the 22.3 minus the 18.5 from the incidence rate in the control group at 7 days, which is 17.9, one gets the following result: the early p value again, the first step, was nominally significant, 0.011, so that we could ask

whether durable clinical efficacy existed. The late p value of 0.071, on its face, might suggest that at 30 days the clinical efficacy was waning. However, this imputed durability p value, taken as a number alone, would suggest that in the PRISM-PLUS trial, in fact, there was significant clinical efficacy that persisted to the 30-day endpoint.

[Slide]

In distinction, in the IMPACT II trial there was a 2.3 percent reduction in the event rates at 48 hours -- difference in the event rates at 48 hours compared with a 1.6 percent reduction at 30 days. The 2.3 percent reduction at 40 days (sic) achieved nominal significance, however, the 30-day did not and when the durability p value was calculated the following results are seen:

The early p value, again 0.015, is significant so the question of durability could be entertained. The late p value, 0.127, suggests again that the durability was not sustained in this particular trial and the durability p value in this case does not change that impression. One interpretation of this would be that this trial alone would not be sufficient to demonstrate persistence of clinical efficacy for this product at 30 days.

[Slide]

In written form then, the methods for establishing the durability of efficacy that I am going through and that

have been proposed by the agency are as follows: From a trial measuring early and late time points, 48 hours and 30 days, you first determine the event rates at those early time points and determine whether clinical efficacy is demonstrated at the early time point, in this case by demonstrating nominal statistical significance. If that significance is not present, obviously examining questions of durability or lack of efficacy would be moot.

However, if that efficacy is demonstrated you next determine if the difference in the event rates at the late time point would be nominally significant if it was applied to the control group of the earlier time point, the 48-hour mark. If nominal significance is retained, this supports the durability of the clinical effect at the later time point.

[Slide]

In conclusion then, the available trials using IIb/IIIa inhibitors have employed primary endpoints between 48 hours and 30 days, and the method of analysis proposed examines the acute effects of IIb/IIIa inhibitors, as well as incorporates information about the durability of the drug's efficacy through 30 days. Thank you.

DR. PACKER: Tom, can I ask you to comment on this proposal?

DR. FLEMING: I believe the rationale behind this is certainly well motivated. If you have an agent that you

fully anticipate to have its signal essentially in the earliest stage of follow-up and you have a strong expectation that evolves after that will reflect neither further benefit nor unintended adverse effects, it is very compelling to say let's look at the signal at the time period where it is most evident, which would in this case be, let's say, at 2-3 days, and then let's explore and ensure that what happens at 30 days at least as showing a consistency of effect.

The issue though that concern me a bit is where this will be most useful is in those settings where a non-trivial additional number of events evolved between the earlier time point and the later time point. We have seen for example, in the PCI setting, if you look at death/MI, the vast majority of the 30-day events are there at 2-3 or by 7 days, even by 2-3 days. On the other hand, in the acute coronary syndrome setting there is more uniformity of occurrence of those events.

So, this method is even more attractive in that latter setting because there is the tremendous opportunity for diluting, and that is really what this analysis is trying to factor out, the diluting that is going to occur if truth really is major signal for the first, let's say, quarter of the events over the 30 days that occur by day 3 and then no signal for the last three-quarters of the events

that occur between day 3 and day 30.

So, I think the rationale is well laid out; it is well motivated. The issues that I would raise though are what are the operating characteristics in terms of what I might call the traditional false-positive or false-negative conclusions?

So, let me at least just pose these and get some response or some thought about this. The first of these would be suppose, in fact -- I mean, is there a risk for a false-positive conclusion? Let's use this example that I was referring to, whether it is death or death/MI. Let's suppose that a quarter of the information that is going to be at 30 days is already in hand at 72 hours. If that is the case -- and let's suppose that the standard error for our estimated difference in death/MI rates is half a percent. That would mean in order to achieve statistical significance at the 7-day point you need to standard errors for the difference. That is about a 1 percent difference.

Well, if you have 4 times the data at 30 days than you do at 7 days, then the standard error of the estimate at 30 days will be twice as large. Well, let's suppose in truth there is a 1 percent difference that exists at 3 days, and let's suppose in truth there is no difference -- the thing we are trying to rule out; the thing we are concerned about -- that there is in truth no difference at 30 days, well,

all you would need to see for this method to give you a positive result is you would need to observe the truth of 1 percent difference, which would be 2 standard errors away from zero and, hence that would be a significant result at 3 days. At 30 days, if you saw just a single standard error away from truth that would be a 1 percent difference, and the chance of seeing 1 standard error is 15 percent. So, this method has a 15 percent false-positive error rate that if in truth there is a 1 percent difference, which is very significant at 3 days, and no difference at 30 days, you are going to get the false-positive impression.

Okay, in addition there are false-negative risks with this. Let's suppose that there is this 1 percent difference, and let's say that 1 percent difference persists -- or, let me even go further, let's say there is 1.5 percent difference, which is 3 standard errors, a p of 0.001 at 3 days, and suppose that 1.5 percent difference truly persists out at 30 days, well, you only have to observe a half a standard error underestimate of that, which will occur with about 30 percent probability, for your observed difference to be less than 1 percent at 30 days which, when imputed back to 7 days, will no longer meet your criterion.

Worse yet, suppose the true difference is 1 percent, just barely hitting 2-sided 0.05, now you have a 50 percent chance, even if in truth that level persisted at 30

days, that the observed level would be less than that.

A third concern is what if in truth, as Reapro authors were trying to establish in their presentation, what if the effect isn't entirely observed in the first 3 days? What if there really is a cumulative effect over time? We will obviously not be taking advantage of that at all so that if, in fact, the difference isn't significant at 3 days but would have been highly significant at 30 days -- too bad, the method fails because you didn't satisfy the initial 3-day condition in the first place and you don't get a labeling indication for 30 days under any of these scenarios. Your labeling indication under any of these scenarios is only for 3 days.

So, I believe the method is well motivated. I fully understand the concern, but there are three or four operating characteristics with the method proposed that leave me very concerned that we could have false-positive or false-negative inferences with this.

DR. SEIGEL: Tom, this method, as any method, has its false-positives and false-negatives, and whether it has more or less, first of all, compared to what? So, this has more false-positives for a 30-day effect than does a 30-day endpoint, but it has far fewer false-positives for a 30-day effect than does just a 48-hour endpoint with no additional controls. So, this is somewhere in between.

I am not advocating, as you know from earlier discussions, 48 hours and this, as you know, reflects a compromise between the greater certainty of a 30-day endpoint with the recognition that that includes more noise and that there is greater efficiency in determining effect early endpoints, and it may or may not have the right characteristics but it is important to see how it compares and make sure we understand how it compares with both of those.

As far as the issue of the false-negatives, one of the principles underlying this approach as opposed to one of the approaches that Doug mentioned, which was the idea that you could look for 48-hour effect and then eyeball whether you thought it was maintained, is that you can actually with any prespecified test power to deal with false-negatives. Yes, it is true that if your drug apparently works at 0.05 in the first 48 hours and has no effect afterwards you have a 50 percent chance of failing, so that gives you 50 percent power, but you can calculate in advance under those assumptions exactly what you need for whatever power you want, which I think is tremendously superior to putting in any criteria which are both arguable at the end because it is not prespecified, and also not possible to power for. So, that is something at least to think about.

Finally, I would say it is probably noting or

thinking a little bit about where this came from and thinking about where the alternatives might be. First of all, where the idea of a compromise came from, it is important to historically note that initially sponsors were told 30 days. Later some sponsors were told 48 hours but you have to measure at least 30 days to make sure -- you know, to take a look at that. And, there are issues of a level playing field that got raised, especially by those sponsors who were told they had to have a 30-day primary endpoint, leading to some discussions about how to reconcile these, and to an approach --

If you accept that it is reasonable to look at a 48-hour primary endpoint and to preserve benefit, and I am not suggesting necessarily that that ought to be accepted, and I won't speak to my personal views at this point, but if you accept that, then the question becomes what is the test to ensure that there is some sort of benefit retained? If you accept that it is desirable to have a test so you don't argue whether it has passed or not and so that you can power for that test, then actually I think the logic for choosing this sort of test was the following:

If the test is simply that you have to have any benefit in the right direction, and that is pretty unsatisfactory, you could lose all your effect but for 1 patient and that means sort of nothing, if the test is that

it has to be significant at 30 days, well, then you are back with the 30-day endpoint which I understand you think we ought to be. If the test is that you have to retain a certain percent of effect, say half of the effect, that is really problematic because that penalizes a very good drug that may both prevent events and also defer some events. So, a drug that has a 6 percent effect and then at 30 days has a 5 percent effect wins, but one that has a 20 percent effect but only has 10 left of that 20, even though it has 10 compared to the drug that had 5, loses. So, a percentage of effect doesn't do it.

So, that sort of led to the idea of having some specific amount of effect, an amount of effect that would be necessary had it appeared at 48 hours.

Let me ask a different question of you, which is if you believe that the data were such that you had some level of comfort that 48-effects were probably retained but you wanted to see data from the trial to ensure that this wasn't an exception to that rule, is this a reasonable approach to look at that in terms of ensuring that you have a prespecified effect size at 30 days, that which would have been significant at 48 hours?

DR. FLEMING: I think this is reasonable but you get what you pay for and I think sponsors ought to be entirely aware that you are paying less and you are getting

less. The basis in terms of the false-positive, I would say is if I believe that 48 or 72 hours is enough the issue is much less complicated. The issue is simple. You just get the data over 48-72 hours. It is clearly a time when we expect the greatest sensitivity. The signal would be large. The sample size will be smaller. That is definitely the easiest solution here.

I am amongst those though that believe that if there is an effect at 48 hours and it is gone by 30 days that it has to be an incredibly non-toxic, convenient and safe intervention to say I have a positive risk/benefit profile if I see a difference that is there at 3 days that is completely gone by as short as 30 days. So, that is motivating me to believe that the benefit conclusion here must be based on data that would establish benefit at 3 days that is convincingly still persistent at some level at 30 days.

Now, what has motivated all of this discussion is the assumption that no difference is going to occur between 3 days and 30 days. The scenarios that I was giving for false-negative conclusions are based on that assumption. What sponsors would need to know, in terms of "you get what you pay for," is that if you power this study to a 30-day endpoint you have a much larger study but, in fact, you have the ability to conclusively establish benefit based on an

achievable difference at 30 days.

If you come in with a much smaller sample size, as this allows, even if you see a 3 standard deviation difference at the 3-day time point -- let's say a 1.5 percent difference, you have a 30 percent chance if the truth is that there is no differential effect between 3 days and 30 days -- a 30 percent chance that the imputation at 30 days back to 3 days is not going to be satisfying.

So, there is truth in advertising here. People have to know that you can mount this trial and the truth could be big effect early; no difference later on. And, there is so much noise in that no difference that it is well within random variation that you are going to see an unlucky result that will not satisfy this criterion, and that is the price you are paying by going with the smaller sample size.

DR. PACKER: Ray, you have the last word.

DR. LIPICKY: I would like to see -- because I think it is the first time that I have ever seen curves of how events go as a function of time on the screen today. I don't know where I have been! As I looked at the shape of the curves, they didn't seem to me to be determined by a single function. It looked like they changed their inflections with time.

So, I am going to make a proposal. Let's say that the time course of events as a function of time after zero

time, in fact, showed that they were best fit by a two exponential fit, something that had a fast time constant and something that had a slow time constant and that, in fact, the drug effect that was present clearly affected the fast time constant and not the slow time constant, and that the fast time constant had a number that was equal to 2 that is its time constant. So, that means by the end of 10 days, whatever that process is that is affected, has expended its total business. If it turned out that way, I would say there is no sense about talking longer than 10 days. You just have to be comfortable that you have truly affected this process.

DR. FLEMING: Let me just allude back --

DR. LIPICKY: And, at 30 days the clinical significance -- to me, clinical significance is getting out of the hospital.

DR. FLEMING: I would like to go back to one of the first comments that I made about this, and that is the properties that we have been discussing depend greatly on what fraction of the 30-day events are already apparent at day 3, and in PCI it is a large fraction. So, the concerns that I am raising are far less in that setting but you are buying much less as well.

Where this method is really motivated is where a fraction of the events where the signal would be expected to occur, i.e., day 3, are a small fraction --

DR. LIPICKY: That is ignoring shape. It is fraction of events. What if the events that you can influence, in fact, are all influenced within 3 days?

DR. SEIGEL: Rob showed us some hazard functions, not simply event rates, and hazard functions are not exactly the slope but closely related to the slope of the event rate for patients still at risk.

DR. LIPICKY: Not at all, it is simply dividing two numbers.

DR. SEIGEL: Okay, and clearly the hazard ratio shows that it isn't the same over the course of time. On the other hand, it doesn't get down to baseline until two or several weeks afterwards. Part of the problem with your argument, Ray, is if there is a separate event that occurs that is a short-term event and you prevent all of those events, you have to distinguish between preventing those events and delaying those events. So, there is a time course without treatment and there is a time course with treatment. And, if you have a type of lesion that might cause a complication by an MI over 3 days, and you give somebody a drug that prevents platelet aggregation for 3 days, you don't know, unless you include days 4, 5, 6 and 7 in your endpoint, whether once that drug wears off those events occur just as often --

DR. LIPICKY: That isn't the result I described.

The result I described is that two exponentials occur. The fast exponential has the entire drug effect and the slow exponential is exactly unchanged.

DR. SEIGEL: Yes, but there is no way you can measure that about a drug.

DR. LIPICKY: I mean, I can certainly fit the curve --

DR. SEIGEL: Because you get a different curve when you treat it with the drug.

DR. LIPICKY: Not time constant-wise.

DR. CALIFF: Well, you can test that. I mean, that is something you could test.

DR. LIPICKY: Yes.

DR. CALIFF: You could estimate that.

DR. LIPICKY: I mean, either I am right or I am wrong. I don't know whether I am right or wrong. I haven't done it.

DR. SEIGEL: But you can't test it by just testing for the first 10 days with that fast curve because you don't delay --

DR. LIPICKY: No, no, no --

DR. CALIFF: This is interagency dialogue.

DR. PACKER: If I could suggest that this discussion is further compounded by the fact that the definition of the components might change because of the

fact that enzymatic monitoring may occur only early and not late, which I would contend is a far greater confounding factor than any confounding factor that anyone has discussed and which I have not yet heard a solution for.

Let me suggest in the interest of time, that we have actually discussed all the questions, whether anyone has noticed or not, and we have provided you with all of the answers.

[Laughter]

There is clearly a substantial number of issues that pertain to the design, the execution and the analysis of positive-controlled trials, including the delineation of the patient population, the delineation of an endpoint, the timing of an endpoint, the selection of the comparator, and you will probably notice that none of those of issues has been resolved today -- none. You have gotten hints today as to how the thought processes may be working, and you have heard some very specific proposals from some very informed and experienced people as to how they might view this. But this is a process that must continue for the future. There are no answers today, but this is a step in the right direction, and I think that the data and the discussions that we have seen today move this process one step forward to try to come up with some way of designing and interpreting positive-controlled trials in this field.

The difficulty, of course, is that we can say anything, design anything, propose anything but databases don't necessarily come out as interpretable as we would like and the proof of the pudding will be when the first positive-controlled trials come in to see whether they, in fact, teach us anything about the efficacy of the drug in the absence of a conventional placebo-controlled trial. Hopefully, we will be able to take a look at the existing data, as well as the data from ongoing and incompletely analyzed trials in order to shed further light on whether these issues can be satisfactorily resolved.

I would like to thank all of the invited experts, our invited guests from industry and all of the committee, and we will consider ourselves adjourned.

[Whereupon, at 4:58 p.m., the proceedings were adjourned.]

- - -

CERTIFICATE

I, **ALICE TOIGO,** the Official Court Reporter for Miller Reporting Company, Inc., hereby certify that I recorded the foregoing proceedings; that the proceedings have been reduced to typewriting by me, or under my direction and that the foregoing transcript is a correct and accurate record of the proceedings to the best of my knowledge, ability and belief.

Alice Toigo

ALICE TOIGO

Lawyer's Notes

8

9

A

B

C

D

E

F

G

H

I