either risk factors or for incident diseases shown in the EPIC study, but it really takes a leap when you hit diabetes, and the same thing has been shown with those categorical studies looking at IFG versus IGT versus type 2 diabetes, the risk is appreciably elevated with IGT and IFG, but it's kind of a 1.1, a 1.2 -- you know, that's the kind of magnitude risk; whereas when you actually develop diabetes, it takes a leap up to two, three, fourfold, which is what the EPIC -- the graph I showed you from the EPIC study.

So it really seems to take a leap upward. Now whether that's related to the glycemia itself or whether, as people go from these pre-diabetic states to diabetic states, they're older, hypertension, obesity, all of those other risk factors are being added on at the same time.

Does that answer in terms of the shape of the graph?

MS. FLEGAL: Well, I guess my question is, once you reach the diabetic level, is there further increase with HbA1c within the diabetic category? Once you reach that level, is the risk the same at all levels? That's my question.

DR. NATHAN: There really haven't been good longitudinal studies that have looked at that, so I don't think we have a sense as to -- once you have diabetes -- you know, glycemia gets worse with age, so you're going to have a whole bunch of other confounding risk factors for CVD as you get into that higher A1c range. The question as to whether lowering glycemia below -- in the sub-diabetic range and lower, has of course never been answered. We're going to hear about the ACCORD study and some of the studies that have looked at lowering glycemia within the diabetic range down to lowish levels, and those results have been summarized here.

They don't appear to give a benefit

for heart disease as yet, and in fact the ACCORD studies suggested that there may be some risk depending on the regimen used, perhaps. I would say that, again, the reason I didn't call this the natural history of anything is of course, all of our patients are being treated much more carefully, much more aggressively for all of their other risk factors, and it is in that setting that we're starting to see in all of these trials lower CVD of end rates.

I mean, it's a good thing. The treatment of CVD has gotten actually much more effective, but it has lowered the -- I mean, it's had huge implications in terms of sample size and power calculations for these trials because the event rates in the placebo treating groups or in the less-intensively treated groups, for example, are considerably lower, and within that therapeutic milieu, it has been so far impossible to demonstrate a benefit of glycemia treatment itself on

cardiovascular disease.

DR. BURMAN: Any other panelist have questions?

DR. TEMPLE: The cardiovascular community has been spoiled by how rapidly the inventions it likes work. You start to see benefits from lowering blood pressure lipids, platelet drugs, within modest number of months. Do you have any thoughts about how long the pure cardiovascular effect of diabetes might take to be either manifested or reversed? And I was struck by your picture of the DCCT study. You don't see any separation until about 12 years.

DR. NATHAN: Right.

DR. TEMPLE: Could that be part of the difficulty, that whatever's going on, it isn't a vascular problem like the others lead to, and, therefore, it's hard to reverse? Any interest or thoughts?

DR. NATHAN: So given time considerations, I didn't put in, especially for this sophisticated group, kind of the different

phases and steps of what's going on for heart disease. I mean, we start off with arthrosclerosis. That takes years to develop and starts probably very, very early, and then you have plaque formation and the breakdown of the plaque and thrombolic phenomenon, and then finally you end up with a clinical event associated with inflammation.

And at any one of these stages, of course, there are probably different mediators of those different stages of cardiovascular disease. In the DCCT, frankly, we were talking about a population that started in an age range, effective age, that was so low that you wouldn't expect them to have clinical events. Now we did measure as well other surrogate measures of arthrosclerosis.

We did carotid IMT measurements. We looked at coronary artery calcification. Those were looked at, of course, at discrete time points. We've done actually I think our

fourth set of carotid IMT measurements, and we can look at that over time. That was evolving, that was getting worse before we saw statistically this increased number of events.

So I think this is all very much predicated on our limited data looking completely at patients -- you know, starting with measuring arthrosclerosis and then following them over a lifetime and long enough to see actually when these signal events occur that cause disease. Having said that, where these various risk factors can be modulated and where they have an effect, a measurable effect, we have all these snapshots in both cardiovascular medicine research as well as in diabetes, and there are just all of these cross-sectional, almost snapshots.

We really have -- in my personal view -- is little understanding of where in this pathogenetic stream of events that you

actually can interfere effectively, and where glycemia would have a beneficial effect. It took us a long time to demonstrate in the DCCT, at least in my opinion, is because it just took us a long time for the patients to reach an age and a duration of diabetes and exposure to risk factors where they got clinical events.

DR. BURMAN: Thank you.

MR. FLEMING: Since Dr. Temple raised this issue, this was one that intrigued me as well. I don't know if we have that slide that we can put back up again, but it looks at this long-term issue with DCCT. With all of these data, obviously, they provide clues. This is a post hoc analysis. I don't know about multiple testing over time, p-values above .0018 most of the time, we would look at with great caution and secondary endpoints with multiple testing, so there are some uncertainties about the conclusiveness of the result, but let's say it's true.

DR. NATHAN: Let me just -- a factual correction. We actually didn't do any analyses over time. The a priori statistical plan here was that we would do any analyses until we had enough events in the placebo group, that we had a chance of seeing, I think it was a 25 or 30 percent reduction, so there were no repeated tests going on here. This was not a post hoc analysis. This was actually done at a discrete point in time based on an a priori test that we did, so we were not looking repeatedly. We were just collecting the data. We only analyzed it when the placebo group --

MR. FLEMING: And that test was set up when the trial was originally designed?

DR. NATHAN: It was set up in 1990 -- it was at least 10 years ago.

MR. FLEMING: So it was as the trial was underway?

DR. NATHAN: As the trial was underway. We only did one analysis in 1993 --

MR. FLEMING: But rather than get too

deep into that issue --

DR. NATHAN: Okay.

MR. FLEMING: There are still issues with that, but what you're saying is helpful. It's still a secondary endpoint. The result is interesting, but the issue I really wanted to get at was one that Dr. Temple raised, because what you're seeing here is a suggestion or an indication of a difference that's long-term that emerges a number of years after the difference in glycemia levels have disappeared; correct?

DR. NATHAN: Correct.

MR. FLEMING: Are there other differences that persisted? As your presentation very eloquently laid out, there are so many confounding risk factors. Are there other differences in these two groups that might explain this beyond the glycemic control?

DR. NATHAN: Several things. Number one is that this is not the only late effect after this A1c between the groups. Again, a 2 percent separation for 6-1/2 years, then

followed by A1cs that were statistically indistinguishable for the next 10 or 12 years, and we've coined this term "metabolic memory" for the microvascular complications, because in fact you continue to see a separation of the retinopathy, nephropathy and neuropathy after the end of the formal study when these A1cs have come together, and so we've demonstrated that even before we showed that. That's number one.

Number two is that -- are there other explanations for this? Was one group more hypertensive? Was one group -- the answer is that every factor we looked at did not explain this, including what you might expect would be the separation in kidney disease, because we in fact had less kidney disease in the conventional treatment group than the intensive treatment group, and when we did the analyses, controlling for the development of micro (inaudible) or kidney disease, these results remained essentially the same.

I mean, it explained a small fraction or a modest fraction of the difference in heart disease, but in fact, the difference in heart disease persisted, even when we control for -- again, all of the variables and the risk factors. Now the number of events here, as you may know, is extremely small, so it limited our ability to do multi-factorial analyses, but those that we were able to do, it didn't explain this finding. It really looked like glycemia.

MR. FLEMING: So the final thought on this, then, is, assuming this is real, it does point out on the setting the importance of very long-term follow-up to really understand the true benefit-to-risk?

DR. NATHAN: Well, especially when you're starting a population that starts in the age range where they're getting arthrosclerosis but not clinical events, and then following them over an average of 18 years until they got to be that age where clinical events were occurring.

DR. BURMAN: And if I could just ask one question. On that slide, when did they become -- differences between the two groups become statistically significantly different?

DR. NATHAN: So this slide -- we, again, we didn't look at any other time -- I shouldn't say that. We looked at one time at the end of the DCCT itself in 1993 when there were numerically a greater number of events in the conventional versus the intensive group. But the numbers were something like a 12 versus 3 or 4, so there was a suggestion, but the number and the event rate was so tiny that we couldn't include anything. That was the first time we looked.

The second time we looked was here, so we didn't do analyses looking at (inaudible) separated, but it was in 1995 when we looked.

DR. BURMAN: Very well. Thank you. Are there any other questions from the panelists? No? Then thank you very much for

the presentation.

I would like to introduce our next speaker, Dr. Robert Ratner.

DR. RATNER: While the slides are coming up, thank you, Mr. Chairman. It's a great honor and pleasure to be here with you.

Although this FDA session is to review cardiovascular disease, my task is to make sure you don't forget microvascular complications of diabetes. As Dr. Nathan described, the original definitions and thresholds for diabetes were determined by the specific microvascular complications, so I don't want to minimize cardiovascular disease -- we can't do that -- I simply want you to remember that all the discussion of cardiovascular complications have to be in the context of what we know and what we are certain about in terms of microvascular complications.

So what are the numbers? This is from the CDC, talking about every day in the

United States -- 4,100 new cases of diabetes. We know that there are 230 amputations for diabetic neuropathy, diabetic foot ulcers, non-healing ulcers and infected ulcers on a daily basis, and that diabetes accounts for the vast majority of non-traumatic amputations.

We know that diabetes is the single largest cause of blindness in the United States, with 55 new cases daily, and it is the number one cause of kidney failure, end stage renal disease (ESRD) in the United States, with 120 cases daily.

Those are the facts that we know about what happens to people with diabetes. And in this morning's New York Times, there's an article that talks about the insidiousness of diabetes and the fact that it is in fact doing silent damage as we go along. These are the things that we know.

We're doing better. It used to be that diabetes was not only the single most

common cause of end stage renal disease, but the only one that was continuously increasing, and what one can see is that in the last 25 years, there's been a remarkable increase in the prevalence of diabetic end stage renal disease, but over the last 5 to 10 years, we're starting to see a leveling off. We're making an impact in terms of end stage renal disease, not soft endpoints like microobunaria (?) but here, end stage renal disease requiring transplant or dialysis, with the incidence rate definitely coming down. So year by year we are getting better. We are clearly making changes.

What about visual impairment? Slowly but surely over time, what we're beginning to see is a fall in the prevalence of diabetic retinopathy. Why is this? Basically because of the studies that Dr. Nathan has presented which have shown the relationships between glycemic control and microvascular complications.

Slowly, gradually, we have improved the level of control in the United States, with the HbA1c levels falling so that in fact we are able to reduce end stage microvascular complications.

When we begin to look at incidence rates, what do we begin to see? These are data from Seattle, from Scott Ramsey's studies in the Group Health Collaborative, looking at what happens to a diabetic population compared to a non-diabetic population in a managed care program. So here, you're looking at almost 9,000 individuals with diabetes compared to 35,000 non-diabetics, and what are the risks that we begin to see?

No question, we see a two- to three-fold increased risk of myocardial infarction and stroke in the folks with diabetes. Here are the absolute numbers of what you see. Soft endpoint, hypertension, about a 1-1/2-fold increased risk.

Now we get into the microvascular complications. We're looking at a threefold increased risk of end stage renal disease, and you begin to look at the comparable numbers for end stage renal disease which is clearly related to glycemia. Foot ulcers, an eightfold relative risk as compared to the non-diabetics. And eye disease, 20-fold increased risk in the individuals with diabetes.

When we begin to look at what diabetes puts people at risk for, clearly cardiovascular disease is there. Please don't forget the microvascular complications as well. If we begin to look at comparable end stage disease from the EDC study in Pittsburgh -- from Trevor Orchard's work, looking over 30 years, you can see renal failure requiring dialysis or transplantation depending upon the cohort from the 1950s, '60s, '70s, '80s, and '90s. You can see the relative risk of renal failure as it occurs

in this population as compared to total coronary artery disease.

Cardiovascular disease needs to be addressed. It has to be addressed because it is what ultimately kills people with diabetes. Let's not forget what leads up to it and causes much of the early morbidity and mortality.

So what's the pathobiology? What is the biologic rationale for thinking that diabetes and glycemia can cause complications? This is a slide from Michael Brownlee's Bantum (?) Award lecture. The highest award given by the American Diabetes Association. Summarizing an enormous amount of work that shows at different levels how glucose can result in abnormalities leading to complications.

With pure hyperglycemia increasing shunting through polyol pathway. With increased levels of metabolites of glucose, an excess in the hexosamine pathway. Later

on, activation of protein (inaudible) pathways, and finally, advanced glycation of end products accumulating within tissues resulting in abnormalities. All of these progressing directly from hyperglycemia.

What Dr. Brownlee has done is to try and put this into a system in which you can understand how glucose could potentially result in the pathobiology of microvascular and macrovascular complications of diabetes, and you can see that he can't solely express it as hyperglycemia. Hyperglycemia clearly is on the background of genetic determinants, and has acute metabolic changes with cumulative long-term changes in macro molecules, but all of this is being influenced by independent accelerating factors, the confounders that Dr. Nathan has described -- hypertension, obesity, dyslipidemia, hypercoaguability -- all of these play on the changes that are already ongoing, the common soil that Dr. Nathan

described.

So there clearly is a pathobiology that could explain the increased prevalence of disease. Now, Dr. Nathan showed you the NHANES and Egyptian data that tried to give a threshold for the definition of diabetes. This is just a bit more recent data from the AusDiab study essentially looking at the same relationship, here with retinopathy. So looking at the population as a whole, you see a very flat and low level prevalence of retinopathy until you get to a HbA1c, somewhere between 5.7 and 6.1 and then it takes off.

If you now take out the group with established diabetes, you still see that threshold phenomenon right around 5.7. When you look at microalbuminuria, a bit more of a slope here in the lower levels of A1c, but again, a clear-cut threshold at approximately 5.8. Now, these are prevalence data, but these are large, large populations that

simply look at the association of glycemia and these specific microvascular complications.

We have to turn to the intervention studies really to be able to make the claims of whether or not this is really causative, and perhaps the best data, as Dr. Nathan described, is coming out of the DCCT. Now I'm not going to show the outcomes data from DCCT, but rather the relationship between the complications and HbA1c. So here you see in a recent publication by John Lachin, the relationship between HbA1c, whether you're in the intensively treated group or you're in the conventionally treated group with diabetic retinopathy, and you can see that regardless of what group you're in, the longer you've had diabetes and the higher your HbA1c has been, the greater the probability of developing diabetic retinopathy.

The question is, is this all time?

Is this all glycemia? What are the contributing factors to the development of these microvascular complications? Again, from the same publication by Lachin, looking at the relationship between glycemic control and these complications -- retinopathy, starting with a single three-step progression going all the way down to laser therapy and macular edema.

Nephropathy, going from microalbuminuria to fixed albuminuria, and neuropathy at five years. And what I want you to concentrate on are the r values and the percent explained by A1c. When you begin to control for all of the other potential confounders, what you begin to see is 95 percent of the effect appears to be related to the HbA1c, to the level of glycemia over time with R-squareds that are shown here.

So in interventional trials we can also draw the relationships between glycemia

and microvascular complications.

Again, as we move into interventional trials trying to prove the relationship, the first, and I think one of the definitive studies of our time is the UKPDS, looking at the relationship in patients with relatively new-onset type 2 diabetes and cumulative microvascular endpoints, with a p-value of .0099, 25 percent relative risk reduction in renal failure or death, vitreous hemorrhage, or photocoagulation by improved glycemic control in the intensive group of this particular study.

As you begin to look at the UKPDS, and I'm sure that Dr. Holman is going to go through this in much greater detail, if you focus exclusively on the microvascular events, what you begin to see is a 12 percent reduction in any diabetes-related endpoint, a 25 percent reduction in microvascular endpoints, breaking it down with a 21 percent

reduction in retinopathy and a 33 percent reduction in albuminuria -- not microalbuminuria, but in albuminuria. So what are the relationships from an epidemiologic standpoint? For every 1 percent decrement in A1c, the UKPDS found a 37 percent decrease in microvascular outcomes. We have to deal with what we know, and we can't ignore it to answer new questions.

Now, we also know that there is a common soil phenomenon here as Dr. Nathan suggested, and we don't treat glycemia in isolation, and one of the most interesting studies that has been published recently is the Steno 2 Trial which asks the question, what if we do everything right? What if we aggressively treat blood pressure, aggressively treat lipids, aggressively anti-coagulate, get people to exercise and eat healthy and stop smoking?

What impact do we have there?

Well, these are data from the microvascular component of Steno 2. I'm not going to address the macrovascular. I'll leave that up to our other speakers. But looking throughout the study, intensive therapy when it came to nephropathy, consistently at four years, eight years, and even after the study was ended, had a significant reduction in nephropathy. Retinopathy, the same -- after the study, that the change becoming a little bit less. And autonomic neuropathy, a greater than 50 percent reduction with this multi-factorial intensive management of diabetes.

So we clearly have evidence that when you begin to approach diabetes as a disease of an individual with multiple confounders, we can clearly reduce microvascular complications. The question really becomes, how do we look at micro and macro at the same time? Well, this is data

from the ADVANCE study which was recently published in the New England Journal and presented at the ADA last month, and they had a very interesting approach, because they started with combined primary outcomes of major macro and microvascular events.

The study design here was to have a sulphonylurea-based intervention versus a non-sulphonylurea-based intervention, and a separation in terms of glycemia. And what you see is that the intensive group had a statistically significant reduction, 10 percent relative risk reduction, in this combined primary outcome. So you treat patients to a HbA1c of less than seven, you get benefit. You clearly get benefit.

Where does the benefit come from? It comes, almost exclusively, from a reduction in major microvascular complications, so that you have a p_value of .015, a 14 percent relative risk reduction, and it's the microvascular complications that

are driving the positive primary outcome in the ADVANCE trial. When you begin to look at the microvascular complications overall, it's statistically significant. New or worsening retinopathy is trending towards a benefit that in fact does not meet statistical criteria. The new or worsening nephropathy, on the other hand, has a statistically significant 21 percent relative risk reduction within the advanced trial.

When you begin to delve even deeper into the renal events, you see a decrease in total renal events, a decrease in new microalbuminuria, which is one of the strongest risk markers for the development of CVD, and a substantial 21 percent risk in new or worsening nephropathy.

So these are the facts that we know. If you look at ADVANCE, 10 percent reduction in combined primary outcomes being driven by predominantly the nephropathic changes with a 21 percent reduction there, no

significant effects on macrovascular events, no significant effects on all cause or cardiovascular mortality, and the changes are consistent throughout the study, no subgroups seem to be different.

So where do we go with this? Here you look at the advanced data broken down by micro and macrovascular disease. The combined endpoint meets statistical power for significance, but the macro does not, and it's driven by the micro. Now the difficulty becomes how do you test for this without adversely affecting that, because we know that interventions that lower glycemia decrease the risk of microvascular complications. Are we going to be able to design studies to look at macrovascular without sacrificing microvascular? That really becomes the dilemma that you're going to have to face.

Let me end with this slide from UKPDS as well. Simply looking at, again, the

ongoing relationship between updated HbA1c and the hazard rate for microvascular versus macrovascular complications, this has been well-reproduced in multiple studies. That as the HbA1c rises, the risk of severe microvascular complications increases. There seems to be a threshold somewhere around six or seven -- nobody really knows where -- that perhaps that's the point of inflection for increased risk, and that if you can get the HbA1c down, you decrease the risk.

The relationship with macrovascular disease, as Dr. Nathan so eloquently showed, is far less steep and far more confounded. Lots of other influences -- insulin resistance, hypercoagulability, blood pressure, lipids -- a whole variety of issues.

How are we going to design a study to look at the relationship here, with this always being kept in mind?

Clearly, one of the ways to do it

would be to look at the HbA1c range way down here. Look at the difference between a group that are controlled to less than six versus a group that's controlled to seven or seven and a half. That's an ethical study. That is a necessary study, and you're going to hear the results of that study shortly. That could be done, theoretically.

Can you look at a patient population comparing the group down here versus a group out here? I would suggest to you that if you need a HbA1c difference between groups of 1.5 percentage points, that the lowest you're going to be able to go in terms of your intervention group is going to be somewhere in the vicinity of 6-1/2, because once you start getting up to mean A1cs above 8, is there an institutional review board in the United States that's going to allow you a 6-, 10-, 12-year exposure of individuals sitting at HbA1cs of 8 and higher?

Let's remember what we know. The relationship between glycemic control and microvascular complications is implicit in the definition of diabetes.

There is clear-cut epidemiologic evidence that as glycemia goes up, there appears to be a threshold -- somewhere in the high fives and low sixes. Interventional trials have definitively shown in both type 1 and type 2 diabetes, that intervention to lower HbA1c, even at the range of seven to nine, significantly reduces microvascular heart events, and there is good pathobiology to suggest why microvascular complications are directly related to glucose.

As you deliberate, I want you to remember not only that diabetes is an important cause of cardiovascular disease, but diabetes is the most common cause of severe microvascular disease as well.

Thank you very much.

DR. BURMAN: Thank you, Dr. Ratner.

This discussion is open now for questions.

Dr. Konstam?

DR. KONSTAM: That was great. You know, maybe you can tell us a little bit about the need for additional diabetic drugs. And the reason I bring it up is because later on, I think we're going to be asking ourselves what level of excess cardiovascular events or cardiovascular mortality we'll feel need to be ruled out if we're interested in cardiovascular safety. And to me, that's not a question that can be addressed in a vacuum; it has to be addressed relative to the potential gain. And you've eloquently indicated that glycemia is related to microvascular events, and we have drugs to reduce glycemia, so I guess it sort of begs the question, what more do we need? How much more do we need from the next drug?

DR. RATNER: Excellent question. If you go back to the early trials of control and complications, the DCCT was aiming to get the HbA1c less than 7 percent. They didn't get

there. They got to 7.2 in an ongoing fashion. If you look at the ACCORD trial, they were aiming to get to a mean of less than 6. They couldn't do it.

And what you begin to see is that the mean HbA1cs in most of the control trials hover somewhere between 7 and 8. Now, part of that is the natural history of the disease. I'm sure Dr. Holman will go through UKPDS showing the updated means, because the A1cs were rising throughout the study, and the limiting factor is that we have to keep adding new medications in. So the question is, why don't the new medications work? Why are they not adequate? And I think that there are multiple different reasons for that.

One potential reason is what has been called clinical inertia. Physicians and patients are reticent to add in new medications until there is true failure, true failure. It's not uncommon in our clinic to

say, Mr. Jones, your HbA1c and your blood sugars are too high, we need to add a new medication. And Mr. Jones says, oh, I just got back from vacation. I know I was eating more. Give me another three months. And that three months turns into a year and a half.

The second, and what I think is an even more important factor is what Phillip Cryer called the limiting factor in the treatment of diabetes, and that is hypoglycemia. All of the therapies that we have traditionally used, most of the therapies that have been in the most recent studies, have as major side effect, hypoglycemia. Now, you're not going to see a whole lot of hypoglycemia if you're starting with individuals at 10 and you're only trying to get them to 8. Although you clearly do see some in the standard treatment groups, and it's really bad when you do.

When you start pushing towards six

and seven, there's less margin for error. There's less for them to fall without becoming symptomatic, so I personally -- and this is solely my belief, is that we need drugs in the treatment arm for diabetes that don't carry with it a risk of hypoglycemia in the near-normal glycemic range. In addition, I would suggest that we need drugs that don't exacerbate obesity, that don't exacerbate hyperlipidemia, that don't exacerbate hypertension, and it would be wonderful if they actually improved cardiovascular disease.

I personally don't believe that diabetes drugs need to be approved solely on the basis of a reduction of cardiovascular disease.

DR. BURMAN: Thank you. Any other questions from the panelists? Yes?

DR. GENUTH: This is really more of a comment. Both you and David Nathan have shown us very persuasive epidemiological relationships

between HbA1c and risk of retinopathy as the classic example, but those are average curves which are the result of looking at sometimes thousands of patients. In reality, that average curve is probably made up of a hundred splayed individual curves.

And so the point I wanted to make is that each patient may actually have his or her own curve and we really don't know what is the lowest HbA1c to aim for in the patient sitting across the desk from us in order to minimize or even eradicate his risk for complications. I realize the FDA has to deal with groups, not with individuals -- but just as you didn't want us to forget microvascular complications, I don't want us to forget that it's the individual patient that we end up treating.

DR. RATNER: I couldn't agree more, Dr. Genuth, and I think that the American Diabetes Association has inappropriately taken a lot of criticism for the table that Dr. Nathan

showed where the goal of the A1c is less than seven, and a lot of people have argued that that's not low enough. Others have argued that it's too low.

What's written in the text, though, is a little bit different. What's written in the text is that you should aim for the lowest HbA1c achievable without unacceptable hypoglycemia. So coming back to the previous question, if we actually have drug therapy that maintained the homeostatic balance between insulin secretion and glucagon secretion and all of the other counter-regulatory hormones so that we could decrease that risk of hypoglycemia, then in fact, we would start going lower and lower.

We can't achieve it safely. And I think that the ACCORD trial and the ADVANCE trial clearly demonstrate that. That's our limiting factor. And frankly, that's why I think we need to be exploring new therapeutic avenues.

DR. BURMAN: Thank you. Dr. Proschan?

MR. PROSCHAN: Given that you've shown that the microvascular events are increasing if you don't control HbA1c, it seems like there's a trade-off. So if a new drug causes MIs but decreases microvascular events -- I mean, some of these microvascular events are more serious than others, and I'm wondering if you have any recommendation about how to consider the seriousness of the microvascular versus macrovascular.

DR. RATNER: I think the dictum most of us follow is first do no harm. And clearly, the microvascular complications are not drug-specific, they are glycemia-specific. So if you have the capability of lowering glycemia with a drug or a collection of therapeutic regimens that don't increase macrovascular disease, that's absolutely appropriate.

I think that when it comes to the cardiovascular complications, those, at least to date, appear to be drug- or perhaps

class-specific. With microvascular, we're just talking about glycemic control. It doesn't matter how you get there. The data have been demonstrated in sulphyonylureas, with metformin, with insulin, so what really matters is getting the glucose down for the microvascular.

DR. BURMAN: Thank you. Dr. Veltri?

MR. VELTRI: That was an excellent presentation, as well as Dr. Nathan. A couple of comments. Obviously, you develop drugs to improve symptoms of diabetes -- polyurate, polyfascia, et cetera -- to improve well-being of patient. And also, some degree then (inaudible) on the microvascular relationship has been clearly established.

Obviously, the macrovascular complications to date have not been established -- indeed potentially, there may be harm, and part of that harm may be related to the fact that so many surrogates, if you go too low and you have an ischemic

substrate, you could have a U-shaped type of phenomenon, if you will.

The questions I have is, number one, are there relationship between the microvascular and the macrovascular? So it could be that a patient population -- and this might actually explain the latency and the affects between microvascular to macrovascular in the DTTC extension.

Is there that relationship? Because clearly, there are relationships among the various microvasculars -- the eye and the kidney.

And secondly, would you think that perhaps a more intensive regimen longer-term, that didn't extend to the DTTC, may actually manifest macrovascular improvement?

DR. RATNER: There are data that look at relationships, and they are not causal, they are solely associative between microvascular complications and macrovascular surrogates, if you will, so that, for example, in the VADT,

Peter Rieven has published work looking at the relationship between stages of diabetic retinopathy, a purely microvascular complication, to coronary calcium scores, and it's curvilinear. As retinopathy goes up, the degree of coronary calcifications goes up. How much this is confounded by time, duration of disease, or level of glycemia, is unclear. Those studies haven't been done.

The clearest relationship is microalbuminurea to cardiovascular disease, and in virtually all studies, the presence of microalbuminurea is a very strong predictor of cardiovascular events, so there may in fact be a link between microvascular and macrovascular. How long that linkage takes is clearly unknown. The suggestion is 12 to 18 years, in DCCT/EDIC, and it becomes difficult in an evolving disease to keep up with the therapeutic changes and still be able to have a clean outcome.

DR. BURMAN: Dr. Goldfine?

DR. GOLDFINE: I'm going to actually ask you just to speculate on something. This is a little bit unfair, but I think that the effect of lowering blood sugar on microvascular complications is absolutely clear, and it's a steep relationship. The relationship may be much more subtle in the cardiovascular end. The other question then also has to do with when are we initiating the intervention, because many patients who have diabetes, by whatever measure you do, already have some established arthrosclerosis, and that the reversal of the phenomenon -- we know that the microvascular disease, you can prevent the development and slow progression, but for an established, calcified, scarred fibrotic plaque, it may be a very difficult time to intervene with existing disease which is already present in many of these people, and there was some interesting data about the importance of early intervention.

And how this might then weigh on to how we should be evaluating this is I think

another important question that you sort of alluded to, and therefore I'd like to push you a little bit on it.

DR. RATNER: Dr. Nathan showed you the data from the diabetes prevention program retinopathy study, which showed there was diabetic retinopathy even at IGT, so we begin to question what is IGT, what is pre-diabetes, what is diabetes? And I think that's a very legitimate discussion to have. The question, though, of whether or not you need to begin intervention at that point for macrovascular disease is almost impossible to answer, however. In the diabetes prevention program, we recruited middle-aged individuals, 50 percent of whom have metabolic syndrome at study entry -- and our cardiovascular event rate, adjudicated cardiovascular event rate, was .08 per 100 patient years. So that's a real problem.

How long is that study going to have to go for the event rate in the control group to get to a point where you have any

chance at all of seeing a benefit? Though intellectually, I believe, starting earlier is better. From a clinical trial standpoint, the statistical power is impossible.

DR. BURMAN: Dr. Parks?

DR. PARKS: Thank you, Dr. Ratner, for your excellent talk. You may have recalled that in that issue of the New England Journal in which the ADVANCE results were published, there was also the results of the ACCORD trial, and the editorial comparing and contrasting those two studies. And earlier, Dr. Nathan had talked about why the intensive arm of ACCORD was stopped early.

My question here is that do we as of yet know about the microvascular complications in the intensively treated arm of ACCORD? And I understand if you cannot answer the question. Perhaps another speaker can.

DR. RATNER: I am not an ACCORD investigator, and so I'm not privy to all of the

data there. Dr. Gerstein is. We'll leave that entirely in his hands. My understanding is that they do not have that data available yet. I certainly have not seen it.

DR. BURMAN: All right. Thank you very much, Dr. Ratner. No other questions? What I'd like to do is have a break and we will now take a 15-minute break. Will the panel members please remember there should be no discussion during the break amongst yourselves or with any member of the audience.

We'll resume at 10:35.

(Recess)

DR. BURMAN: Take your seats, if you would. We'll get started in a minute. Please take your seats.

Why don't we get started? We will now proceed with further guest speakers' presentations. Dr. Thomas Fleming will be discussing and evaluating the benefit and risk of type 2 diabetes statistical considerations.

Dr. Fleming?

DR. FLEMING: Thank you. What I'd like to do is, as just noted, focus on some of the statistical issues that arise as we're looking for reliable evaluations of benefit-to-risk in type 2 diabetes. And the main focus of what I want to talk about will be on evaluation of safety issues, but I'd like to bridge the presentations that we've had by briefly talking a bit more about surrogate endpoints and validation of surrogate endpoints. So when we're looking specifically at biomarkers in diabetes, we have some very good ones.

We've heard a lot about HbA1c, clearly establishes biologic activity, and as we've discussed in some depth already today, there's considerable evidence for its reliability in understanding microvascular complication effects -- retinopathy, neuropathy, nephropathy -- much more controversy and uncertainty about effects on macrovascular complications.

And so these effects on HbA1c are not necessarily giving us the reliable understanding of the overall clinical efficacy. And everything is always benefit-to-risk, and so the effects as well on HbA1c may not be able to reliably predict what the global safety or risk profile will be for the intervention.

And so as we look at surrogates, what are some of the things that we think about that influence our sense about their reliability? And I'll talk about a couple of specific issues. One is understanding that with any disease process, there are multiple pathways through which the disease process causally influences the clinically tangible important outcomes or consequences for patients, and if in fact the surrogate endpoint lies in one of these pathways, we could get either false negative conclusions or false positive conclusions by relying only on information about the effect on the

biomarker.

But even in a setting such as type 2 diabetes, where we've heard considerable evidence about the ability of HbA1c to capture, in essence, a principal causal pathway, there still are important issues about what is the magnitude of the effect on that biomarker; that is, the targeted level to optimize the effect of the intervention on the clinical outcomes? What is an adequate level of effect to predict clinical benefit? What is maybe an over-effect? And also, what is the duration of that effect that's needed?

In addition to the fact that the intervention can have the intended effects on the causal pathways, interventions can have mechanisms of action that are independent of the disease process, and in fact, this explains very often why an intervention's effect on a biomarker may not reliably predict what its ultimate effect is on the

clinical endpoint because of these unintended mechanisms of action.

The literature is full of examples of where surrogates have gone awry, and some of the recent examples that we've already heard discussion about -- in the ACCORD trial, the strategy for more intensive glucose control against a 7 to 7.9 target did in fact show a reduction did in fact achieve the intended reduction of HbA1c, but suggested at least an increase in mortality.

This type of phenomenon has existed in the past in other settings. With erythropoietin in renal and oncology settings, getting more proper standardization or normalization of hemoglobin to more ideal levels hasn't yielded the intended reduction, but in fact an increase in mortality.

Quickly to review this, the goal here in end stage renal disease in patients with high risk of cardiac complications was to provide a more complete normalization of

hematocrit levels to reduce the risk of death and MI, where standard dose Epogen was yielding hematocrit levels of 30 percent, and so treating to a higher dose of Epogen was the experimental arm to achieve a more complete normalization of hematocrit.

And what we saw in the trial, looking at the relationship between the hematocrit level and the percent deaths is as the hematocrit level went down in the control arm, the death rate was higher. And in the intervention arm, the same phenomenon was seen -- as hematocrit levels were lower, the death rate was higher, such that looking at the pool of data, for every 10 point increase in hematocrit, one had a 30 percent reduction in the risk of death.

Then looking at the patient distributions in the standard arm, most patients were in the 30 to 33 range, and with a more intensive does of Epogen, one was able to achieve a standardized level of

hematocrit. So it would seem logical to then conclude that because models would show that in both the control arm and the intervention arm, as you achieve more standardization, you achieve lower levels of death -- and the experimental arm did in fact render patients at a more standard level than the standard arm -- one would expect, then, that there should have been a reduction in death rate. Well, in fact, there was rather than a 25 percent reduction in death rate, there was a 30 percent increase in death rate.

And on our data monitoring committee on which I served, when we did the interim analysis at half the planned events, when we had 366 patients with the primary endpoint where the expectation or the hope was that the high dose, achieving a more standardized hematocrit or hemoglobin level, should have given about 40 fewer deaths and MIs, a 25 percent reduction, there was in fact almost 40 increased deaths and MIs, or a

30 percent increase, which was statistically significant even adjusting for the multiple testing aspect allowed one to rule out even the most trivial improvement in what was intended, which was a reduction in death, a reduction in death and MI.

Well, as the data were explored, it looks like this may well have been mediated through an unintended increase in thrombosis.

There are a number of other examples that we've had discussion about where, even though we've achieved the intended reduction in HbA1c with troglitazone, separate independent risks, serious hepatic risks -- and we've got examples where even though we've achieved the intended effects on biomarkers, the very endpoints that we were trying to improve have been worsened with the addition of torcetrapid to atorvastatin, we not only achieve reductions in LDL, but the increase in HDL, and yet as we know, we had an

unexpected increase in death, in cardiovascular death, stroke and MI, and the examples that we have discussed already, rosiglitazone and muraglitazar, while we are able to achieve reductions in HbA1c with muraglitazar, a suggested increase in death, stroke, and MI, rosiglitazone suggested increase in MI.

In each of these settings, the issue of particular concern is while these interventions are affecting surrogates such as HbA1c, providing benefit maybe on some of the clinical component outcome, such as microvascular complications, could there be unintended mechanisms not captured by the effects on the surrogate that give us a net effect on the true clinical endpoint that are adverse or not consistent with what you'd expect them to be just by looking at the effect on the surrogate?

So I'd like to spend a couple of minutes talking about the issue of validation

of surrogates, beginning with the definition of a valid surrogate. A valid surrogate arises in a setting where the effect on the intervention on the clinical endpoints, so the totality of the effect on the clinical endpoint, is reliably predicted by the effect of the intervention on the surrogate.

And so to illustrate this validation process, let's look in the setting that, for example, was studied in the ACCORD trial, where the clinical endpoint was cardiovascular death, MI and stroke, so lambda represents the rate of the clinical endpoint, and the intervention, the control, Z equals zero and the intervention active, experimental Z equal one. So in a classical proportional hazards model, one is modeling the effect of intervention on the endpoint of cardiovascular death, stroke, and MI, and broken down into simpler terms, lambda-0(t)is the clinical endpoint rate in the control arm. Lambda-1(t) is that rate in the

experimental arm, and one is hoping that the experimental rate is reduced from the control rate by some constant multiple of the proportional hazards model.

So in the ACCORD trial, the intention or the hope was through intensive glucose control compared to more standard glucose control targets, that we would be able to detect a 15 percent relative reduction in the rate of cardiovascular death, stroke, and MI, and to have 89 percent power to do so with a traditional 2.5 percent false positive error rate requires a trail to have a very large, 1,540 events, which even with a trial of 5 to 6 years follow-up, would be 10,000 patients.

Clearly, if we can understand the relationship of interventions with clinical endpoints in trials that are much shorter and smaller, it is one of the major potential benefits of using surrogate endpoints.

So what are some of the principal

criteria we have to consider to determine whether a surrogate is valid? Well, first of all, it needs to be correlated with the clinical outcome, so if HbA1c is the biomarker, clearly it is necessary that it be correlated with the clinical outcomes of interest, but a correlate does not a surrogate make.

The far more complicated and critical criterion is that the surrogate needs to fully capture the net effect of the intervention on the clinical outcome, and to look at how one can get evidence regarding whether that is true -- let's consider this same setting as an ACCORD where the primary endpoint is cardiovascular death, stroke, and MI, wherein the intervention is looking at is the control, standard glucose control against intensive glucose control.

But now let's not only look at how intervention affects the outcome rate, if we want to look to see whether HbA1c could be a

valid surrogate for how intervention is affecting the clinical endpoint, we model not only the treatment arm, but also the HbA1c at a given time. And if in fact HbA1c is in fact a valid surrogate fully capturing how the intervention affects this clinical outcome rate, then in this given model, gamma will be non-zero, because in fact we already have validated that HbA1c is correlated with the clinical outcome. But the key issue is, if in fact HbA1c at any given time is fully capturing how the intervention is affecting the clinical outcome, then beta should be near zero.

In other words, once you've factored in how the treatment affects HbA1c, there's no residual or additional effect of treatment on the clinical outcome. This is the type of evidence that we would be looking at to get further validation that the biomarker is capturing accurately how treatment is in fact influencing the effect

on clinical outcomes.

The reality, though, is in essence, what we would then do is look to see whether beta is much smaller than alpha -- is in fact there evidence that the essence of the effect is being captured by the biomarker? Or the proportion of the net effect explained by the surrogate might be 1-beta/alpha. One of the problems is, beta/alpha is much more variable than alpha, and so it takes multiple times, more data, to estimate beta/alpha than it does alpha.

So in other words, to validate a surrogate endpoint, you need clinical studies that are powered to assess what the effect of the intervention is on the true clinical endpoint, and you need many of them to be able to then -- to start having enough data to determine whether the biomarker is a valid surrogate.

The concept that we might validate a surrogate endpoint in a phase 2 trial and

use it in phase 3 is only valid if your phase 2 trial is many times larger than your phase 3, which is in fact not the case.

So meta-analyses are required. The other issue is, even if in this particular analysis -- let's say with HbA1c, it does appear that the effect of an intervention on the clinical endpoint is fully captured because beta is near zero, you're only looking at the net effect. And to illustrate this, suppose that an intervention provides a 15 percent reduction in the rate of major clinical endpoints or major clinical events, and suppose that's exactly the level of effect that would be predicted by what the effect is on HbA1c.

It doesn't allow you to conclude that the only way that the intervention effected the outcome was mediated through its effect on HbA1c. There may have been undetected positive effects through other mechanisms and undetected negative effects.

And if these counterbalance in their magnitude, then the analysis that's looking at whether you're fully capturing the net effect will give you in fact an answer that, yes, you are. And yet the entire effect isn't specifically mediated through HbA1c, and that's important because new interventions that come along may have different balances in these mechanisms than the intervention that was studied that was used to "validate" the biomarker.

Now, this type of analysis can also be used not only to get information about whether the mechanism to achieve benefit was mediated through the surrogate. It can also be used to get some clues about whether when there's evidence of harm, was that harm mediated through a defined outcome? So in the ACCORD trial, where -- let's say, now the endpoint -- let lambda be death, the death rate. So in the ACCORD trial, the intensive glucose management -- the intensive control

against standard control suggested an increase in death rate -- in this case, either the alpha was positive, was a number greater than one; i.e., evidence that intensive glucose control may have had a harmful effect on mortality -- one of the questions is was that in fact mediated through an increase in hypoglycemic events?

So we can use the same kind of analysis to get clues about that. Specifically, we look not only at the effect of the intensive versus standard glucose control, the effect on mortality, but we also factor in the hypoglycemic status at a given point in time. And if in fact the effects of this intervention on mortality is in fact mediated through the hypoglycemic episodes, then beta would be near zero again, or if beta on the other hand is near alpha, then you would be saying the actual mechanism through which this intervention led to the mortality increase was not related to the

effect on hypoglycemic events.

One however has to be very cautious about interpreting this, particularly in settings where beta is near alpha; i.e., where you get the apparent conclusion that the negative effect on mortality was not mediated through hypoglycemic events. That in fact might be a false negative conclusion if you're mismodeling the specific nature of the hypoglycemic covariate here. So if you're modeling it as whether at a given time you are hypoglycemic, if in fact what you're missing is the level of hypoglycemia or the duration of hypoglycemia, then it may be that the treatment effect that was negative on mortality may have in part been mediated through hypoglycemia, but you're missing it with the modeling.

It's also possible that you'd be getting a false negative conclusion here if this variable is highly variable. So for example, in an anti-hypertensive setting

where the outcome is stroke and you're looking at blood pressure, when we've done these kinds of analyses, even though the effect of an intervention on stroke is undoubtedly substantially mediated through effects on blood pressure, these types of analyses may not reflect that, and that's because blood pressure is such a variable measure that the measure is not capturing the true blood pressure that someone has, or the true mechanism. You're going to get an attenuation of effects.

So as you use these kinds of analyses, they're giving you clues -- at best clues about the mechanism through which you achieve the effect.

Ultimately, to validate a surrogate endpoint requires a comprehensive understanding of the causal pathways in disease process as well as the intended and unintended effects of the intervention, and it's very difficult to have a comprehensive

understanding of the unintended effects, they're generally unintentional, frequently unrecognized and undocumented. Ultimately, the best evidence for validation of a surrogate comes from meta-analyses of clinical trials data.

So hypothetically, this would be the kind of evidence -- for example, if we were trying to look at the degree to which effects on HbA1c could be a valid surrogate of, let's say, macrovascular complications -- cardiovascular death, stroke, and MI. Suppose we do a large number of studies, suppose about 20 separate studies -- and in each study we look at what is the treatment versus control difference in effects on HbA1c, and we plot it against the treatment versus control hazard ratio or effects on the clinical endpoint of cardiovascular death, stroke, and MI.

This would be an ideal setting for validating the surrogate. In settings where

there is no net effect on HbA1c, there's essentially no effect on cardiovascular death, stroke, and MI. When you have a moderate effect, you have a moderate reduction. When you have a substantial effect, you have a substantial reduction.

These kinds of data would provide the best evidence to validate a surrogate.

In type 2 diabetes, when we're looking at validating HbA1c, these kinds of analyses can be done, and as is well-motivated by the discussion we've already had today, validating HbA1c could be in fact successfully achieved for certain classes of endpoints but not for others, and in fact, it's important when you're looking at a biomarker, in a setting where there are multiple clinical endpoints that are related to the disease process or the treatment for that disease that are very clinically important, it is important to be looking at whether the biomarker is valid for all

aspects of these specific outcomes.

An example of this are in anti-hypertensives. On June 15, 2005, the FDA Cardio-Renal Advisory Committee met to look and to probe to what extent has blood pressure now been validated for an array of clinical outcomes. And specifically, the data that were provided for this validation involved randomized comparative trials of more than 500,000 patients.

And the totality of these data allowed us to look at the extent to which blood pressure lowering was a valid surrogate for these clinical endpoints separately across classes of agents. Low dose diuretics, beta blockers, ace inhibitors, calcium channel blockers, ARBs, and that's one of the important issues, is when you're validating a surrogate, technically speaking, you need to validate it for each separate class of agents, because the unintended mechanisms that can affect the reliability of

the prediction of the effect on the clinical endpoint based on the biomarker, can differ across those indications.

And what was found with these data was that blood pressure gave a very good prediction of the actual effect on stroke across all of these -- nearly in all instances across these agents -- moderately well for MI and cardiovascular disease, not quite so well for mortality, and not well for heart failure.

And to give just one illustration of this, of the kind of evidence that was provided, it was looking at the extent to which systolic blood pressure differences were predicting effects on cardiovascular events. And so in this particular display across the X axis is the degree of effect in reducing systolic blood pressure. The further to the right, the better. The Y axis was giving the clinical outcome, the relative risk for cardiovascular events, hopefully

looking at reduced values being more positive effects -- and the wide array of trials that are listed here were used to look at the relationship, and this is a slide from Henry Black's presentation of that advisory committee.

And what we see is a definite relationship here with blood pressure, that as interventions achieve a better effect in reducing systolic blood pressure, you are seeing a reduction in the rate of cardiovascular events, although with some diminishing returns. More is not necessarily better. So kind of a common theme that we're seeing potentially here with HbA1c and that we've seen with ESAs, erythropoietin stimulating agents.

What I'd like to do now is to move to some specific issues or challenges we're going to have as we look at evaluation of safety. When we're assessing safety issues, everything is benefit-to-risk, and so the

stronger or more compelling the evidence we have for efficacy, the more resilience we have to what level of confidence or certainty we have in safety. There are many issues, there are many examples that have arisen in recent times where we have interventions that have substantial effects on symptoms, or interventions that have effects on biomarkers for more substantive clinical outcomes.

And yet in those settings, there is a lack of resilience to what the overall benefit-to-risk would be if these interventions actually had an unintended negative effect on measures of irreversible morbidity or mortality, and these are all examples in recent times where these situations arose.

The COX-2 inhibitors provide important analgesic effects and reduce GI ulceration rates relative to non-selective NSAIDs in patients with rheumatoid arthritis and osteoarthritis. Long acting

beta-agonists provide reduction in symptoms of severe asthma. Anti-psychotics have been important for patients with schizophrenia.

And in the setting where effects have been shown on biomarkers, in agents that have been approved with biomarkers, rosiglitazone and erythropoietin provide beneficial effects respectively on HbA1c or overall hemoglobin levels. But in each of these settings, there are concerns about what true benefit-to-risk would be because of potential or established negative effects on measures of irreversible morbidity or mortality.

So increased risk of cardiovascular death, stroke, and MI that are occurring at rates of 1.5 to 2 could substantially alter the benefit-to-risk of these interventions, or increased effects on mortality with erythropoietin of 10 to 15 percent, potentially even as much as a fourfold increase in mortality in the long acting

beta-agonists -- also are settings where these unintended effects substantially alter the overall benefit-to-risk profile.

The primary goal is to be able to identify effective interventions that are safe. And in these settings where efficacy is for a symptom, or efficacy is on a biomarker or a surrogate endpoint for clinical outcome, there's more concern that the safety issues could be sufficiently substantial to alter the true benefit-to-risk, and long-term and rare outcomes can be very influential. The goal in these types of settings then would be to rule out that you have unacceptable increases in safety risks in order to be assured of having favorable benefit-to-risk. And very quickly, there are numbers of sources that we have for such safety information.

Passive and active surveillance and large-scale randomized clinical trials provide us both pre- and post-marketing.

Most often, the surveillance approaches are post-marketing, and these can be useful for both surveillance of new safety signals and exploration of existing signals.

Very quickly, the post-marketing Adverse Event Reporting System with a voluntary submission of MedWatch forms does provide us a timely way of getting signal detection or hypothesis generation, but by its voluntary or passive nature, it provides a less reliable aspect; hence, this approach is really only particularly effective for detecting risks that are large relative risks that particularly have a close temper relationship with the intervention. In essence, while they are timely and uniform, we lack having denominators and numerators.

And so a somewhat more rigorous approach would be through active surveillance, large link databases or through a perspective pharmaco-vigilance program that is looking at prospective cohorts. And while

this approach does give us numerators and denominators, it still is weakened by the fact that the data comes from a non-randomized setting, and there are other issues of sensitivity and specificity that are non-optimal.

So for these particular reasons, these approaches are particularly effective when you're trying to detect, or when you are detecting, very large relative risks. So with Tysabri for progressive multifocal leukoencephalopathy, for PML, when this should be a one in million rate, when it's occurring in studies at one in a thousand, that's a thousand-fold relative increase. Or with the rotavirus vaccine, with intussusceptions, more than a tenfold relative increase. Here is where the post-marketing surveillance systems are very effective in being able to detect safety risks.

On the other hand, in many of these

other settings, these safety risks that we're talking about on cardiovascular death, stroke, and MI, a 1.5 to twofold increase, or increases in mortality of 10 to 15 percent, or even up to a fourfold increase, these levels of relative risk are much more difficult to reliably discern what is a true treatment-induced risk just from selection factors as to who received the intervention and who didn't.

Randomization, having a randomized trial, systematically removes these imbalances. Patient and caregivers don't start and stop therapies at random. And so if we're only using data from active surveillance or passive surveillance, there's a tremendous risk of confounding what is the true treatment effect from these selection factors.

Also, safety assessments should include among other evaluations ITT evaluations, Intention To Treat evaluations,

that require the ability to have a time 0 cohort. Assessment of risk over a specified time interval is key even if the intervention is stopped earlier in time.

So for example, with the COX-2, there's been some concern that even if you stop Vioxx earlier in time, the overall effect of the intervention, adverse effect on cardiovascular death, stroke, and MI, might in fact be something that's only realized later in time -- unless you have a time 0 cohort following people beyond the time they discontinue therapy, you're not going to be able to assess that outcome.

Risk can't be assumed to be independent of duration of exposure. So in breast cancer, if you're giving Adrimycin, it's perfectly fine until you get 450 cumulative dose, after which, major cardiovascular risks occur. And from data that we've seen today, benefit safety issues are in fact a combination of beneficial and

negative mechanisms. And so it may well be that when you're looking at the long-term impact of a type 2 diabetes agent on safety outcomes, those could be very different from short term.

Having 10,000 people followed for six months, whereas it's 5,000 person years of follow-up, isn't necessarily giving you the same insight as having 1/10th of that 1,000 people followed for 10 times as long, 5 years, and again, this kind of insight was apparent from Dr. Nathan's presentation, that relative effects, both safety and efficacy effects long-term, may not be represented by short-term.

Having a -- whether it's randomized or not, prospective cohort is key for being able to have enhanced sensitivity and specificity being able to adjudicate events, being able to retain increased retention and being able to achieve high levels of adherence. You can't rule out a safety risk

if people have substandard adherence to what it is that you would be typically using in practice.

So how big would these trials typically have to be? Well, suppose you are looking at -- in the setting of the PAX-2 inhibitors, where there's a 1 percent rate or a 10/1,000 rate, if you wanted to rule out a tripling, it would take 2,000 person years -- or with the long-acting beta-agonists, where it's a 1 event per thousand 1,000 person years to rule out a tripling would then take 10 times the sample size or 20,000 person years. These analyses of person years are based on the assumption that you'd want 90 percent power to rule out this increase -- if in fact there is no increase -- while having only a 2.5 percent false positive conclusion -- only a 2.5 percent of risk for saying there's no increased risk when there really is at this level.

But allowing 20 increased cardiovascular deaths, strokes, and MIs in a COX-2 inhibitor setting would be an inadequate assessment of safety. Even a smaller increase such as an increase of five events per 1,000 person years would be important; hence, you would need 20,000 person years in this setting. In type 2 diabetes, where you might have a 20/1,000 baseline rate, to rule out this excess of five events per 1,000 person years could take 40,000. And so as was seen in the ACCORD trial, if you're following people for five years, you might need a sample size of 8,000 to 10,000 to be able to rule out this 25 percent relative increase, or this increase of 5 events per 1,000 person years.

Let me just quickly walk you through one specific trial where this type of assessment was done. And this study that I'm going to look at with you is in the setting of COX-2 inhibitors. And specifically, this

is a trial, a safety study that is currently underway in patients with osteoarthritis and rheumatoid arthritis, looking at the pain medications Celecoxib against ibuprofen and naproxen, and the specific interest here is to determine whether or not one can rule out that the COX-2 inhibitor has an unacceptable increase in the rate of cardiovascular death, stroke, and MI.

So this is a trial being conducted in a setting where ample evidence exists for concern about an increased risk, but where the thought is that Celecoxib might in fact, if the dose is being given recommended, might in fact not share the same excess risks seen with other COX-2 inhibitors.

And so to give you a sense of how this study is being constructed, I'll focus in particular on the COX-2 as the experimental and naproxen as the control. And so lambda-0(t) represents the rate of cardiovascular death, stroke, and MI in

Naproxen, and the question is, is Celecoxib in fact -- is the rate of Celecoxib not an unacceptably large increase over the rate on Naproxen?

And what's been defined as the level that has to be ruled out is a one-third increase. And so the hypothesis that one would want to be able to rule out is a one-third increase in the setting where there is no increase, so where beta = 0. So the study is designed in a manner such that when in fact there is no increase, you'd have 90 percent power to rule out a one-third increase, where, however if in fact there is a one-third increase, you would get a false positive conclusion of safety only 2.5 percent of the time.

To achieve that, the study has to be of sufficient size and duration for 508 patients to experience the event of cardiovascular death, stroke, and MI. So if in fact this trial of 508 events, or a 20,000

person trial, is conducted, how do we analyze the results?

What I'm showing here along this axis is the relative rate on Celecoxib, the COX-2 against Naproxen, for the end point of cardiovascular death, MI and stroke, so a favorable result for Celecoxib would be one where its relative rate is lower than Naproxen. An unfavorable result is off to the right here, where its rate would be unacceptably high.

The null hypothesis, or the hypothesis that has to be ruled out in order to establish adequate safety, is that the rate on Celecoxib is at least 1/3 higher than the rate on Naproxen. With 508 events, one will be able to in fact rule out a 1/3 increase if in fact you see no more than a 12 percent increase.

So the least favorable result, this result or anything to the left, would rule out a 1/3 increase, and essentially after

much discussion, based on the analgesic benefits of Celecoxib, based on its reduction in the rate of GI ulceration, it was determined that it would be acceptable as long as it doesn't yield, essentially, three additional cardiovascular death, strokes, or MIs per 1,000 person years, and the result will be positive if the estimate is no more than one excess cardiovascular death, stroke, and MI per 1,000 person years.

Now, how do you interpret the results? If in fact the result is no more than a 12 percent increase or better, then one rules out the margin of 33 percent and would conclude that you have in essence non-inferiority, or ruling out an unacceptable increase.

Conversely, if you have at least a 19 percent increase or anything worse than that, you'd actually be ruling out a quality, establishing that you're inferior.

In a result here in between, you'd

be neither inferior nor establishing non-inferiority, and of course if the result is highly favorable, where there's a 16 percent relative decrease in the risk of cardiovascular death, stroke, and MI, the confidence interval would rule out equality, so even though your goal was to at least be able to rule out an increase, you could in fact establish that you're superior on that particular outcome.

Now, some insight, added insight, would occur here by considering a hypothetical case. What if the trial was done not with 508 events, but with 1,000 events? So you actually followed these patients such that 1,000 of them had an outcome of cardiovascular death, stroke, and MI, and suppose you had an estimated 15 percent increase.

Then this trial would successfully rule out unacceptable harm, would establish non-inferiority while proving you're

inferior. Now, you have to be a statistician, I suppose, to find that okay. I'm okay with that. This is a setting where this trial would establish non-inferiority while proving you're inferior. Okay?

But it's semantics. What does it mean when you're establishing non-inferiority? There was a trial done not long ago by a sponsor in this type 2 diabetes setting where these kinds of results occurred, and when this occurred, the sponsor said, this allows us to conclude that our experimental therapy is at least as good as the active comparator -- because we've established non-inferiority, we can conclude we're at least as good as the active comparator.

Well, that's not the conclusion that you can make by establishing non-inferiority. Clearly, they're not at least as good as. They're inferior. To state you're at least as good as, you'd have

to be superior. Superiority rules out any level of being worse. This is what you'd have to see in order to state you're at least as good as. Essentially here, what you're establishing is that you're not unacceptably worse than, so that's why I have no problem with non-inferiority, yet proving inferiority.

Non-inferiority simply means that you don't have an unacceptable increase in harm, even though you may have an increase in harm. It's not an unacceptable increase. And that points out why this margin is critical. This needs to be the smallest excess, which if real, wouldn't be acceptable. If in fact a 10 percent excess would be unacceptable, then a 33 percent margin is an inadequate establishment of safety.

Now, I want to spend a couple minutes on a critically important issue. Properly conducting these safety studies to

rule out unacceptable excess requires very careful attention to performance standards, to ensuring you have high quality conduct.

The first of these is you need to have timely enrollment. This is especially important if it's decided that these safety studies can be done in a post-marketing setting. If you have evidence of efficacy, let's say on microvascular complications, you're going to market a product for some considerable period of time, while you then, in a post-marketing setting, conduct a study to ensure that the overall net benefit-to-risk is favorable -- if it takes an extended period of time to enroll the trial, you're not getting from a public health perspective an adequately timely result.

The target population of ineligibility rates need to be such that you're addressing settings where the excess risk is most plausible. But at the same

time, you need to be sure you're getting a sufficient event rate, because the essence of those trials, the power of the trials, isn't specifically the numbers of patients and duration of follow-up, it's the numbers of events. And so the higher the risk population, the more events. But again, it has to be a risk population relevant to where you're concerned about excess safety risk.

Retention is key in order to be able to maintain integrity of randomization. So if we look at the RECORD trial, for example, the RECORD trial was intended to go after a group that had 11 percent risk rate per year, and got only a 3 percent rate per year. It was intended to have only 2 percent loss to follow-up, but had 50 percent relative higher rates of loss to follow-up. These two consequences impact the timeliness and reliability.

The ADOPT trial had a lower enrollment that was intended, had a lower

risk level or event rate that was intended, had higher levels of loss to follow-up than was intended and had a withdrawal rate of nearly 40 percent.

The consequences of all of these impact the timeliness and reliability of the results. So for example, the FDA in their May 29, 1999 letter of approval for rosiglitazone indicated that a long-term four-year trial was needed, including an assessment of long-term cardiovascular risk that was to be provided by the ADOPT trial.

And yet this study was only published in December of '06, so it came 7-1/2 years later in time, and even at that time provided only 68 MIs across three groups, so roughly 45 per pair-wise comparison they weren't adjudicated.

And so issues that were violating these key principles had a big impact on the timeliness and reliability of the results, but adherence and cross-ins are particularly

critical. So let me just go back to the previous slide for the moment. High levels of adherence and lack of cross-ins is critical in a safety study where you're trying to rule out an excess risk.

Suppose for example that Celecoxib really does provide at least a one-third increase in the risk of cardiovascular death, stroke, and MI. Well, if the adherence to Celecoxib is substandard, is less than it would be in a real world setting, you're not doing a true test of whether Celecoxib is giving an unacceptable safety risk. Or if the Naproxen patients are crossing in to Celecoxib, then you may be diluting what that excess risk is, and that diluting could take a true scenario where you have an unacceptable safety risk and give you the false sense that you're not getting an excess safety risk.

So as a consequence, adherence is critical. My view is adherence should match

the best real-world level achievable. I don't want 100 percent adherence if that's not going to be seen in the real world, but I would want best real-world level of adherence, achievable level of adherence. It must at least match the adherence also seen in prior trials that gave rise to the safety signal.

Cross-ins need to be addressed in multiple fashions. The first is through careful screening. So for example, in the Celecoxib/Naproxen trial, we don't need to enroll all patients. We should enroll those patients who have true equipoise. If you think you want Celecoxib, or if in fact you think you have no interest in taking Celecoxib, that's fine, proceed as you wish.

But for those patients that truly have equipoise and are willing to either be randomized and remain on Celecoxib long-term, or to be randomized to a non-Celecoxib and not cross in, those are the patients that

should be entered. So careful screening is critical.

Careful educating of caregivers and patients is critical so that patients understand the nature of the design and why such cross-ins or adherence are critical to the ability to interpret. Then, as these studies are conducted, they need to be monitored. They need to be monitored for these standards.

So for example, in this precision trial that I've been showing you, which is a 20,000 person trial to be enrolled, the target enrollment is a 30-month enrollment period. The rate of events target is 2 percent. Minimally acceptable levels have to be established, 1.5 to 1.75 percent. High levels of adherence targets have been set.

Cross-in levels, a 2.5 percent cross-in target has been established where it would be unacceptable if it were more than 10 percent. Loss to follow-up, retention

rate standards have been set, where a 2 percent loss to follow-up rate is the target. Greater than 5 percent would be unacceptable. Careful monitoring then during the course of this trial of these standards needs to be done, and this is exactly what's happening now in this precision trial.

So in conclusion, there are multiple instances where surrogate endpoints have been used. They've been used for accelerated approval as with Tysabri, they've been used for full regulatory approval as with ESAs, rosiglitazone. In these types of settings, we get -- by virtue of the use of the surrogate, we get less reliable evidence about efficacy and less reliable evidence about safety. And everything is benefit-to-risk.

Ultimately, the stronger the efficacy evidence, the greater resilience you have to uncertainties about safety. So if we're using biomarkers as the way to assess

benefit, then we are less resilient to what might be an unacceptable safety risk.

And in development of interventions in diabetes, it is important to be efficient here, and biomarkers provide us an enhanced way to be efficient, certainly giving us a more timely result, but it's key to have reliability as well as timeliness in assessments of both safety and efficacy.

And while timeliness could potentially give us choices in a quicker way, ultimately we can't compromise reliability because in essence what patients really care about isn't just a choice, it's an informed choice.

Thanks.

DR. BURMAN: Thank you, Dr. Fleming.

Yes, Dr. Holmboe? Did you have a question? Yes.

DR. HOLMBOE: You talked a little bit about prospective cohorts, and I just wonder if you could give us your feelings on one form of a

prospective cohort, and that's registries, where you have the capability of collecting some information, prospectively from the get-go that may be adventurous down the road, that as you point out in large databases while they could be very helpful, you're stuck with what's in them. You know, you can't obviously add stuff.

So I would just be curious, because this keeps coming up, not only just in this context, but I know in other meetings you've been at, this idea of how do we follow this stuff along when you have these difficult risk/benefit ratios. And you highlighted a number of the things that have really challenged us. So I'd like to hear your thoughts on that.

DR. FLEMING: Sure. Registries are very important. Having large cohorts, particularly in settings where they are prospectively assessed, which would be more like an active surveillance system, where you have a greater ability to achieve high levels of

sensitivity and specificity and adjudication, are valuable. I see them particularly valuable for being able to describe natural history. What happens to patients? What is the overall event rate? What are the covariates that are predictive of that event rate? How are patients managed?

So for all of those purposes -- by the way, some of those purposes are very valuable to planning clinical trials, because they give you a sense of what event rates would be. They're valuable for counseling patients for prognosis. They're valuable for helping us understand where there's an unmet need. The weakness of those is providing us information about causal effects of interventions and outcomes, so if we're looking at very large relative risks, it works.

It worked for Tysabri with 1,000 relative risk. It worked for in its inception at a relative risk of 10. But in

so many settings, what we care about clinically are relative risks that might be a one-third increase, and to be able to discern what's causally a treatment-induced effect from selection factors is extraordinarily limited.

DR. BURMAN: Dr. Konstam?

DR. KONSTAM: Thanks, Tom. Two questions. One is, I just wonder if you could give us some insight into the sensitivity of the upper confidence boundary to the number of events. So taking the example that you had of the 508 events -- ruling out a 33 percent increase, what would be the comparable number of events for -- let's say ruling out a 50 percent increase? And then I have a second question.

DR. FLEMING: Sure. So essentially generally as you double the difference that you're allowing, you would have one-fourth the number of events required, and that's doubling on a log scale, so if you take the log of .33, at .50, if the log (inaudible) twice, then it

would take one-fourth the number of events.

So it's very tempting to define those margins to be 50 percent, 70 percent, et cetera.

DR. KONSTAM: No, that's fine, but I'm just kind of trying to ask, because I think this is going to be relevant to sort of judging how well we're doing today based on the current approaches to program development, so you're saying that a quarter of 508 would yield you a upper confidence limit --

DR. FLEMING: So just to be real specific --

DR. KONSTAM: Right.

DR. FLEMING: If you were trying to rule out a one-third increase, it would take 508 events. If you're trying to rule out a 50 percent increase, it would take 256 events. If you tried to rule out a doubling, it takes only 88 events. So if we have 88 events and we're not seeing an excess, basically we're in a position to rule out a doubling. If you have 15

events and you haven't established an excess, it's a classic example of absence of evidence isn't evidence of absence; i.e., when we don't have a lot of events, concluding that we're fine is an absence of evidence scenario which isn't evidence of absence, and that's where we are predominantly when we have sources of information with 5 events, 20 events, 15 events.

DR. KONSTAM: That leads me to my next question, because I guess it's not an uncommon practice, and I think we're sort of being asked about this practice today of looking at the point estimate of whatever set of data we have today and if the point estimate is on the okay side of -- is in the right direction or not in the wrong direction, we might say, okay, we're good. But if it's in the wrong direction, then we've got to do a specific safety study. And I won't even ask you to comment on that because I'll bet you'll say it's irrational, but maybe you do think it's rational.

DR. FLEMING: Should I just -- what

you've already said is very rational. It's very important. What you're talking about here is what is my best sense of truth, and that's the point estimate, but ultimately, the reliability of that point estimate matters greatly, so it's not just what it is but what is the confidence interval, what can you rule out. So just to follow up on your point, if we have an intervention that we think actually could provide a somewhat favorable effect on cardiovascular death, stroke, and MI, you can rule out that it provides an unfavorable level using a rigorous margin without a large sample size.

I think there's a misconception that non-inferiority -- this is non-inferiority here. You're trying to rule out an unacceptable safety risk, it requires huge sample sizes. No, it doesn't. Not in a setting where you have an intervention that could be slightly favorable. Now, it might be, and this is pure speculation on my part,

that the six-month or one-year effect of an anti-diabetic intervention could have a somewhat unfavorable effect on relative risk, but it could be over five years somewhat favorable as you in fact start seeing beneficial effects.

Maybe there are multiple mechanism, some unintended negative effects early, but overridden by long-term effects that are eventually seen with glucose control. So if you do a longer-term five-year follow-up trial and you actually have a slightly favorable relative risk like .9, you're not going to be able to power that trial for superiority, but you can power that trial to rule out a 30 percent increase without an inordinately large sample size.

DR. KONSTAM: I guess what I was going to come to is, the alpha that we assign to the assessments I guess has an arbitrariness to --

DR. FLEMING: Yes.

DR. KONSTAM: As does, therefore, how