structure?"

Comments?

DR. DIEPPE: Well, just to open things up, Steve, awfully short, in my view.

DR. LIN: I want to go back to the endpoint Question Number 4 a little bit. Dr. Elashoff said that the problem here is that we have joint-space narrowing, and we're going to the second phase of this thing, but the withdrawal criteria is based on joint-space narrowing. So how do you deal with dropouts based on that? And if we just simply assign failures on pain or function, those patients who dropped out based on joint-space narrowing, in a way I think that that's just validating the hypothesis, or that's what we're trying to prove. So there is a complication there.

DR. JOHNSON: I think most people thought it was a bad idea to drop them out based on the joint space.

DR. LIN: Right, but then the question is, if a patient really has joint-space narrowing so severe, are you going to keep this patient on the trial?

DR. JOHNSON: Yes.

DR. ABRAMSON: Assuming that they are fully informed about their choices.

MS. MALONE: Exactly.

DR. ABRAMSON: There would be a drug that's

been approved for structure, and they'll have to decide to continue in the study or not. But they shouldn't be withdrawn --

DR. JOHNSON: And as a consequence, some will drop out, I'm sure.

DR. ABRAMSON: Right.

DR. JOHNSON: But at least the blind will be maintained.

I think Maxime's question is interesting. Nobody wanted to answer it. But if something dramatically affected osteophytes, you'd have to wonder what's going on. You'd probably have to go through a whole argument about digging up whatever epidemiology existed, and I don't know how it would compare to joint-space narrowing epidemiology. Maybe Marc does.

DR. HOCHBERG: Well, there's an interesting controversy which has come up in a journal called Preventive Medicine. There are a number of epidemiologic studies that appear to have shown a protective effect of hormone replacement therapy on radiographic changes consistent with osteoarthritis at the hip and at the knee, and there's one paper published which, in a prospective longitudinal study, showed that women who were taking hormone replacement therapy were more likely to develop symptomatic osteoarthritis. This has actually been

confirmed in a second, independent study from a different country using a similar-type longitudinal design and large database. So here you've got the difference between an x-ray osteophyte, which appears to be decreased with HRT, versus symptomatic therapy.

So it's entirely possible this issue that Maxime brings up, which is, if you retard the growth of osteophytes or you resorb the osteophytes with remodeling, that the symptoms may get worse. But clearly they should be measured.

DR. DIEPPE: And if I could just add that Ken Brandt's already provided some insight into that by quoting the previous study showing that when surgeons cut off osteophytes, people did badly.

DR. ABRAMSON: Yes, Ken?

DR. BRANDT: And further data from the University of Chicago on operative specimens measuring medial and lateral instability after the shaving of osteophytes added about 2 to 4 degrees total on both medial and lateral instability. So there is stabilizing to a degree, and you increase instability by removing them.

DR. ABRAMSON: Getting back to the duration, for structure I guess 1 year is what we're starting with, and sort of finding our way based on the sensitivity of our measures.

DR. DOUGADOS: It's obvious that the duration of the study is related to the sensitivity and the reproducibility of the technique, but we have to be very careful, because on the studies we have conducted, during the first year, if you conduct a 5-year or a 3-year placebo-controlled trial and focus on the placebo, during the first year there is a huge degradation, and then the degradation is less important, probably because the people willing to participate in the 3- or 4-year placebo-controlled trial are in an active phase of the disease, which is completely different than the usual population, and my personal explanation is because they have an inflammatory process, and this inflammatory process is responsible for an episode of chondrolysis, which is able to be demonstrated within 1 year.

So now with the tools we have for assessing the structure for it, I do feel that it is possible to demonstrate a statistically significant difference of an active drug within 1 year. The question is related to the generalizability of the results. So yes, it's possible after 1 year, but if we focus on the specific patient with an active inflammatory course of the disease, is it a reality? Now, in daily practice, I am not sure, and because of that, I should recommend tc look at the second year or the third year.

DR. ABRAMSON: Dr. Dieppe?

DR. DIEPPE: Well, I completely agree with Maxime. Our data is exactly the same in our studies, that rate of progression was faster in the first year of observation than all subsequent years, and like Maxime, I think that's an issue related to ascertainment of patients. So if patients are ascertained for drug studies in the same way as Maxime and I have ascertained them for our longitudinal studies, I think it's too short, and I think we should not be driven by the technology in deciding how long this thing is.

I think that's a real issue with all this chatter about MRI. You know, we'll have it down to 3 weeks soon, and what the hell does that mean to a patient in a disease course of 20 years if we can power a study at 3 weeks? I think it's nonsense.

DR. JOHNSON: Yes, the question was meant to be extra-technology. This is a question about what duration, in principle, should be desired, given that the disease lasts 20, 30, or 40 years. I'm sure, in fact, if you can do it in a year, I can do it in 6 months with twice the number of patients, and with MRI, I can probably do it in a month's time. That's not the issue. The question is what should -- now, I don't know if this ascertainment business -- I mean, that might apply to any decision that you're

ever going to make, I suppose, about duration necessary for a trial.

I guess my concern is are there fundamentals -- and I think I read somewhere or I heard somewhere that there were in osteoporosis, for instance, where they do 2- or 3-year trials, and there's something about the cyclical or the course of osteoporosis in the second year that, from a scientific point of view, you could argue that 1 year is invalid and you need 2 or 3 years. Now, I don't see anything like that in osteoarthritis, but if there were such a beast, we should attend to it.

DR. DOUGADOS: It has been proposed to look at the changes in structure only taking as a baseline value -- not the baseline value, but the data obtained at 1 year -- as an example, such as in osteoporosis -- to conduct a 3-year placebo-controlled trial, to forget the first year, and then to look at the changes of the structure only during the second and third year. That has been proposed. It is not my personal opinion. I think it's possible to look at the baseline as the true baseline value.

DR. ABRAMSON: Dr. Brandt?

DR. BRANDT: Yes, but remember there's also evidence from at least a couple of places to suggest that the rate of narrowing, the annual rate of narrowing, will depend on the joint-space width at the outset, and in those

where it is relatively normal -- 4.5 or 5 millimeters -- at baseline, the rate may be slower initially than it is if the rate is 2.5 millimeters at baseline on this. And we don't quite know where we are on this, but there's a suggestion that that's the case. And it's most rapid when you start with no joint space.

(Laughter.)

DR. JOHNSON: But I guess the question is -- I guess it's always involved with the generalizability of things, but assuming you rev up the most sensitive subset you can find and you can do it in 6 months, should that be considered sufficient in duration, or should it be a year? Should it be 2 years? I mean, the 1-year call was relatively arbitrary. I think we use the same in rheumatoid arthritis.

DR. ABRAMSON: Maxime?

DR. DOUGADOS: I think that from the scientific point of view, 5 years or 10 years is greater than 1 year or 6 months. The question after that will be in the next section, the missing information, because the number of dropouts will interfere with the results. But I think for our recommendation, to say at least 1 year seems at this stage reasonable.

DR. JOHNSON: Don't the Europeans suggest 2 years?

DR. DOUGADOS: I know.

DR. DIEPPE: I would go with the European suggestion, being a European. I would go for 2 years minimum.

DR. DOUGADOS: Since I am also European, I will agree with Paul.

(Laughter.)

DR. ABRAMSON: Okay. Then what about the symptom demonstration duration? What do we hear from the Europeans on that? Is it more than 2 years?

DR. ELASHOFF: It seems to me that the issue is what symptoms are you talking about. Are you talking about the symptom of joint replacement or are you talking about just sort of generally speaking whether you feel okay today in talking about that sort of thing?

DR. ABRAMSON: Do you have a clarification on what you're intending here?

DR. JOHNSON: Yes. It's never been joint replacement. I mean, I was excited about joint replacement and got royally shot down a couple of years ago. I mean, ideally it's an interesting endpoint, but it's confounded by so many non-medical factors. The Europeans seem to finesse this whole issue, as I read their document. They just say, "Go ahead, you can use structure as an endpoint, as long as you supply evidence for surrogacy in your

package." So they don't address the issue.

I think this issue bears on what Bill Schwieterman was talking about this morning. You know, if you've got a Phase IV program that's going to go for 5 or 10 years, it's bound to fail. If you've got one that's going to go for 2 or 3 weeks, it's not going to be adequate. So there's probably some optimal duration. But that's from approval time, not from starting the trial, which in most cases may well be ongoing at approval time for structure.

DR. ABRAMSON: Maxime?

DR. DOUGADOS: I would like to emphasize your comment concerning what will we measure in these Phase IV studies, because one proposition is to forget the date of the surgery because it is controversial, sometimes it's related to the salary of the surgeon and not to the stature of the patients. But by analogy with the domain of cardiology, they have proposed criteria for heart failure. That is, indicate not surgery for cardiac transplantation, but the time at which the patient fulfills the criteria as an indication for surgery.

So in the field of osteoarthritis, that will be a good combination between symptoms and structure. If we agree to find something that is a composite index, this patient should take benefit of surgery at this time and

thereafter you can take this, and it will be a composite index between total articular replacement and symptoms. You see what I mean? That is, not the surgery, but the indication for surgery.

DR. ABRAMSON: Marc Hochberg?

DR. HOCHBERG: Well, I think you can assess symptoms at the time that you're doing the Phase III trial for structure modification, where, as is described in the European document, you can have an agent which demonstrates structure modification and symptom improvement or one that just demonstrates structure modification. I think I would go back to the issue of in Phase IV to do a long-term registry without a placebo group as opposed to continuing on to look for long-term symptomatic changes which don't occur within the first 2 years that one is doing the trial to look for structure modification.

DR. ABRAMSON: Dr. Brandt?

DR. BRANDT: Maxime, I'm not sure I followed you, but if you were proposing something like an index for surgery based on Lequesne score and an x-ray score or so, I don't buy it, because I think that as we discussed before and Ms. Malone brought up, levels of pain or function or so have different meanings to each patient, and some patients with a WOMAC score of 14 are perfectly happy to go with a WOMAC score of 14 -- in fact, they may not even be patients

-- and others with a WOMAC score of 9 may be absolutely falling apart and don't want to go on another day that way. I think it's very tough to make an index of those kinds of things.

DR. DOUGADOS: Yes, but I didn't say that this index should be focused on the WOMAC. I said it might be a composite index, taking into account the capacity of the patient to cope with the osteoarthritis, the level of activity, depression, pain, functional impairment, concomitant therapy, and not only an absolute value, it should be perhaps an area under the curve of the last 6 months or something like that.

DR. BRANDT: I think that doctors, in very simple terms, shouldn't tell patients when they need a joint replacement. Patients should tell doctors when they need a joint replacement, in a sense, in terms of the amount of difficulty they're having on a daily basis. It's very difficult to mandate a procedure like that based on any of these mixes with it, because each patient filters those things differently.

DR. DOUGADOS: I don't mean that this patient should go and see the surgeon. I say in the clinical trial, the way to analyze the patient -- that is a possibility. If you don't want to call these indications for surgery, you can call these failures, and that's

nothing to do with the individual indication for surgery. I didn't say that this patient, if he/she fulfills these criteria, should go to see the surgeon. I say in the analysis we can consider this patient as a failure.

DR. BRANDT: Yes, but that failure is best defined, I think, in patients' terms rather than our terms.

DR. ABRAMSON: I guess one of the other issues, to come back to what Dr. Hochberg raised, and it may not be clear yet either among the group, the committee, or the agency, the demonstration of symptom improvement is a question that's a little murky in the sense that if we took the European view perhaps that structure improvement and no worsening is an indication, and then you might need some long-term follow-ups to make sure that the patients, through a registry or something, don't do less well in some way, is that sufficient? Or are we still uncertain whether eventually the radiographs are only surrogates and eventually true validation for this new drug would have to include symptomatic improvement? Those are two very different decisions in looking at this question.

DR. JOHNSON: In that regard, Marc, spin me a scenario. If you did a registry of 1,000 patients, how would you know at the end of 10 years that that drug had offered a clinical benefit? What would be your comparison? Some historic control or --

DR. HOCHBERG: Well, you could enroll in the registry all the patients that complete the trial. Now, some of the patients who originally were randomized to placebo -- again, I haven't sat down to design this, but let's say some of those who were originally randomized to placebo will choose not to go on this therapy when it becomes available, because they'll say, well, there hasn't been any long-term experience with it, and I don't want to take something that hasn't been on the market for 10 years. It's a possibility. Some of the people will go on the therapy, some of the people who were randomized to the active drug in the trial will eventually discontinue it, some will continue on it, and then if you enroll those people in a long-term registry, you may have some observational data at the end, which, granted, won't be as good as if you did a 5- or 10-year placebo-controlled trial from which you might be able to make some inferences about long-term outcomes.

But it seems to me that if the patients are going to deteriorate in terms of symptoms, they probably will do so within the time that you're doing the study to look at structure modification. Similarly, they should probably improve in symptoms if it has any symptomatic effect within that time. I mean, for straight symptomatic drugs, we study 6 to 12 weeks, sometimes 24 weeks, and now

maybe out to a year.

DR. JOHNSON: I don't know how long these drugs are going to take to kick in, but it's been argued that they may not kick in for 18 months or 2 years. So fundamentally if you didn't continue some randomization scheme, you would have to rely on risk adjusting in the standard epidemiology sense to draw your conclusions.

DR. ABRAMSON: Janet?

DR. ELASHOFF: Well, assuming that you have reasonable follow-up, you could always recompare the ones who had been assigned to the placebo group with those who had been assigned to the treatment group and see which ones are doing better a long time out, because if you really had a big advantage from that structural change, the ones who had been assigned to the drug group should still be doing better, irrespective of what either one had been on at the time. So in some sense your long-term stuff is based on initial intent to treat. Now, whether you can actually follow up people in a reasonable way there so that that makes sense, I don't know.

DR. ABRAMSON: Dr. Witter?

DR. WITTER: Just taking a little different spin and just a question that I'm not aware of something like this, but kind of getting at what Maxime and Dr. Brandt were discussing, is there such a simple question to

help us figure out what some of these endpoints may be, like a global that the patient would say, "I think I need surgery, yes or no"? I mean, has anything like that ever been -- I have not seen it, I'm just --

DR. BRANDT: There's a good reason you haven't seen it.

(Laughter.)

DR. ABRAMSON: Let's move on, then, to the next page on the analysis, and with respect to multiplicity, "To preserve trial-wide alpha -- that is, risk of a positive result of no more than 5 percent when many analyses are done -- should the alpha be distributed in the scenarios below?"

And I'd like, I guess, to ask Dr. Elashoff to lead off on this, please.

DR. ELASHOFF: Okay. I have one minor comment, and that is, on Number 2, I don't think you'd be using last value carried forward, if that's what that means.

Since there are a variety of different ways that one can think of adjusting for multiplicity, I wouldn't say that we should immediately be thinking that the way it should or will be done is to divide up alpha among a variety of variables. The whole issue of what are possible ways to divide up alpha is somewhat of a statistical one, and I think people should bring back that

kind of thing later on. But the whole thing of which are really the primary variables and which are the secondary ones or to what extent they can be ranked as to importance, I think is reasonable to discuss here.

DR. ABRAMSON: Other comments?

Is it worth going through -- and perhaps, Janet, I'd ask you to help on this -- going through and responding to the specific Questions 1, 2, and 3 there? Can you lead on it?

DR. ELASHOFF: Well, I don't know very much about these variables, because arthritis is not my area. I would say that to some extent how you would group them and how you would prioritize them does depend on how they're correlated. If there is some group that tends to really hang together, then you might either pick one of those as an important primary variable or a composite of those as an important primary variable. But in terms of which ones you might rank first, I can't comment.

DR. ABRAMSON: Dr. Dougados?

DR. DOUGADOS: I think this point is important to clarify. If you consider the joint-space width is the most important variable to consider, we have to focus the primary outcome measure on the changes in the joint-space width during the study by treatment group. But you have also to keep in mind that there are at least seven or eight

possibilities to present the results in terms of the changes in the absolute value in millimeter, the percentage of change, the percentage of change per year, the percentage of change between the final minus the baseline, the percentage of patients with relevant progression, as we discussed this morning, with the relevant progression based on the SDD technique, the smallest detectable difference, all based on the clinical relevant technique, as I have presented, or you can say that every year you will calculate the percentage of patients who will progress, using the life table analysis, and the event is defined by the relevant progression. So you see there are eight possibilities, even if you are focusing on one single measurement that is a change in joint-space width in millimeter.

That is the reason why the discussion this morning was important also, to see what is clinically relevant, presenting the results in terms of as a continuous variable. That is, the change in millimeter per year, taking into account the dropouts, or the percentage of patients with relevant progression. We know that we will lose in terms of statistical power if we switch from the continuous variable to the dichotomous variable, but perhaps we will win in terms of clinical relevance, and if we are focusing today on a potential drug which is able to

demonstrate a symptomatic effect during the time of the development and to get the registration for the structure, personally I should strongly recommend to present it on an individual basis. That is, the percentage of progression, yes or no.

DR. ABRAMSON: Yes?

DR. LIN: I think this multiplicity issue here, you have to first better define the hypotheses you want to test. You know, if we go back to the previous guidance that you have four endpoints and you want three of them to win, the way I heard was previously we didn't know how to deal with that, so we insisted on four out of four, and that's when we -- not to have any kind of alpha adjustment. But the field of multiplicity in the statistical community, they have developed -- we have quite a bit of understanding of how to deal with that problem, three out of four. So that's something we can deal with right now.

But going back to the list of questions here, I'd say that you really have to structure your hypotheses, and if you structure your hypotheses in a logical way, you don't have to make an alpha adjustment. I mean, I'm talking about the statistician will know the stage-wise testing. You can have a test at alpha level on joint-space narrowing, and if you have that endpoint -- and we're talking about going to a Phase III/Phase IV to evaluate

symptoms -- if you structure right the test on symptom analysis, you don't have to make an alpha adjustment. Similarly, item 3, the pain analysis. All these can be dealt with if you get your hypotheses structured in a very logical way.

So I just wanted to bring that up.

DR. ABRAMSON: Any other comments? Yes, Dr. Anderson?

DR. ANDERSON: I'd just like to comment and agree with what Dr. Dougados said, that defining an outcome on an individual basis, having sort of responder or non-responder on the whole complex of outcomes, would really be very desirable, and obviously, looking at the set of outcomes, it's extremely difficult. I mean, it was done in RA, but it took about 10 years, didn't it? I don't know whether there's any work being done to get toward something like that in OA yet.

DR. ABRAMSON: All right. Question Number 3 under here is pain analysis and function analysis. Can I ask Drs. Johnson and Witter to just give a little clarification?

DR. JOHNSON: Yes. I think it was the first version of this document, we separated them. We've got a pain claim and a function claim, and there was a lot of debate back and forth about the entanglement of these

concepts, and I think that view held the day in the end, so we put it back together. But they're still measured separately, and there are drugs that sponsors think may work better on one than the other, so it's not fair to say that they automatically should be equally weighed.

Maybe this will reduce to a non-question. I mean, if you do wonderful on function, don't change pain at all, that strikes me as actually bona fide. Or vice versa. So that was the spirit behind the question.

DR. ABRAMSON: Maxime?

DR. DOUGADOS: Two comments. The first one is, I can understand the concept of the differentiation between the domain of pain and the domain of functional impairment, but, unfortunately, the tools we are using at this time in the field of osteoarthritis show that every time we are looking at the correlation existing between the results in terms of pain and the results in terms of functional impairment, they are very closely related, even though I understand that the concept is different.

The second thing is just to remind that one of the questions is how to combine, how to weight, and that is one of the objectives of the responder criteria initiative. That is, to propose a set of criteria which will be a composite index, taking into account the information coming from the three main domains that have been previously

selected by other societies, such as pain, functional impairment, and global assessment. So that will be a possibility to give a result in terms of an individual basis -- again, responder, yes/no -- and taking into account pain and function. Probably the same level in terms of weighting system.

DR. JOHNSON: Will you have indices that incorporate structure also?

DR. DOUGADOS: At this time, we have decided only to propose a set of criteria on symptoms, and symptoms means with efficacy symptomatic parameters, and we do not take at this time, but we will discuss probably in further studies, the possibility to add not only the symptoms in terms of safety and also the symptoms in terms of structure, because the best would be a composite index with symptoms, structure, and safety. But we don't have this available at this time.

DR. JOHNSON: I think we're partially a victim of our prior empiricism, essentially. I mean, of the functional measures, the Lequesne and the WOMAC, they actually are multi-dimensional themselves and have pain, function, and stiffness. I'm gathering that the sense of these prior meetings, such as OMERACT and so on, was that the representation of pain within those indices may not be a full representation of pain, and I'm presuming that's why

everybody -- 95 percent -- wanted to have a 10-centimeter pain scale as a part of all this.

So there's a certain redundancy already, I guess, in the two measures we have, which is, I think, your point. We still have an analytic problem as to how to combine them, though.

DR. HOCHBERG: I'm sorry, I don't know if I agree with that in terms of at least the pain scale measurement within the WOMAC. I mean, granted that within the Lequesne everything is aggregated together, although Maxime has developed some ways of statistically disaggregating it, but within the WOMAC there is in fact not a total WOMAC score. I mean, according to Nick Bellamy, that's not valid. One has to look at the five pain questions separately from the 17 function questions and the two stiffness questions, and you then can ask people how they would combine them together for their individual self, but there's no way of aggregating them together across 100 patients the same way.

So I think there's a little redundancy in the guidance document about using a VAS pain scale and using the WOMAC to measure pain. That seems to me like you're measuring it twice.

But to get back to what was said earlier, for the structure-modifying drugs, if they want to go in for a

structure claim alone, then clearly that's the primary outcome variable, and you want to look at symptoms to make sure they don't deteriorate. If they want to go in for a structure and symptom claim, then there has to be some way of using all three of those as the primary outcome variable rather than having symptoms as the secondary, if that's the claim that they're proposing when they go into initial negotiations with you.

DR. ABRAMSON: Dr. Dieppe?

DR. DIEPPE: A couple of comments, really. The first is that I think most of the studies show that pain is one of the biggest determinants of disability, so they are inextricably linked. But I just wanted to make the point you have to be just a bit careful here about what pain you're measuring, because if you confine pain to the indexed joint, then you're missing a trick, because, of course, it's global pain that correlates best with disability because of what's going on in other joints and other aspects of the patients. So again, there's room for getting into a muddle here by not defining what sort of pain you're measuring, whether it's indexed joint pain or whether it's total body pain.

The other point I'd make is just to say that in relationship to what Marc was saying with Nick Bellamy's studies, we've been doing some qualitative research with OA

patients which suggests that most individuals regard function as much more important to them than pain. Even though pain is one of the main determinants of that function, function is what they care about most. I don't find that surprising, but maybe we should just bear that in mind, too.

DR. JOHNSON: So I gather there remains something that is captured by that pain -- and we meant a global pain assessment here -- that the pain chunk of the WOMAC doesn't capture.

DR. DIEPPE: I think that's my point, yes.

DR. ABRAMSON: Okay. Number 4, rescue medication use analysis, I think we've addressed at this point. Is there another piece of that that anyone wants to comment on?

(No response.)

DR. ABRAMSON: And Phase III analysis and Phase IV analysis, I'm not quite sure if you want additional discussion on that piece.

DR. JOHNSON: That's redundant, essentially. I mean, I think fundamentally, I think logically, if you do a Phase III approval analysis on structure and you win, and you do a Phase IV validation analysis on symptoms and you win, and you don't adjust, you've inflated your error. But it may be that it's really going to be our call in the

sense that we would probably want to take that risk, essentially, or at least the important thing from our point of view is probably going to be the validation study so that presumably the product can stay on the market.

DR. ABRAMSON: So what would happen if a product met structural parameters, but on the validation study there was no clear improvement of symptoms? That product, at least as it's now being thought about, would not continue on the market?

DR. JOHNSON: That's right. There is some sort of regulatory way of withdrawing it.

DR. ABRAMSON: So that would be different, then, from the European approach to registration of such products, or not?

DR. DOUGADOS: At this time, no, because the European guidelines, it states that at this stage the European agency will not approve a drug with only structural benefit. The responder has to provide some information concerning the symptomatic improvement, either with their drug or based on the research of empirical data.

DR. ABRAMSON: Okay. So let's go to the missing information and dropouts.

I'm sorry. Dr. Anderson?

DR. ANDERSON: I just wanted to ask a question about this Number 5, the Phase III and Phase IV analyses.

You're talking in terms of sharing the alpha between those, were you? I mean, if there's a .045 significance on the structure and then 2 years later there was .045 on the symptoms, would that be okay, or would that be too close to the wind, or would the combination not be satisfactory?

DR. JOHNSON: Well, I guess the way you share it depends on how correlated you think the two endpoints are, and since we don't know that, I guess we can't say ahead of time.

DR. ANDERSON: But if you've planned to do it this way from the very beginning, I don't know that -- can you say something about that, Janet?

DR. ELASHOFF: Well, it's also conditional. You never get to Phase IV unless you pass Phase III. So it's not the standard kind of alpha-sharing thing, and I don't see a real problem with it here, but you'd have to sit down and work out what are really the consequences of that kind of plan.

DR. LIN: That's precisely what I was saying earlier. If you plan your hypothesis in a logical order, then if you win on joint-space narrowing and then, going into the Phase IV, do your pain or function at that point, provided that you have passed joint-space narrowing -- and you can test at .05 again, and you will not be incurring more than -- overall, you will not be incurring more than 5

percent error.

So Kent, you were saying that if you win on .05 on joint-space narrowing and then, going into Phase IV, you win again on the .05 level, that your alpha is over 5 percent. That's not necessarily true.

DR. ABRAMSON: Let's go on, then. The missing information, "Best way to analyze, assuming dropouts are informative?" Would someone on the panel like to comment?

DR. HYDE: Actually, I have just a comment on the last one. I guess one thing to think of is, say you do a 2-year study and you find you improve joint-space narrowing, but, gee, we worked well for the clinical symptoms, too; I guess we're done. If you don't, then you go on to the Phase IV. So there's really essentially an interim analysis implicit in that, and there you might get into trouble.

DR. ABRAMSON: Dr. Elashoff?

DR. ELASHOFF: It seems to me with respect to dropouts, as a statistician, I would say the answer to Number 2 should always be yes, if you could possibly do it, because then you lower the missing information. But back to Number 1, questions about how you deal with missing information depend on how much information is missing and what you think are the mechanisms. As to item C, there are lots of different techniques, and I don't think that how

best to analyze it is something that could be answered briefly at a committee meeting.

DR. ABRAMSON: Other comments?

DR. DOUGADOS: If you want, I can show you, just to emphasize the importance of the -- in a 3-year trial with placebo, we had 225 patients, including the ITT population, and only 138 in the completer population, and if you compare the patients in the placebo who dropped out during the study versus the patients who completed the 3-year study, they were completely different with regard both to the baseline characteristics and the rate of progression in terms of the primary outcome measure. In terms of baseline characteristics, they had more severe disease, with a joint-space width that was lower than the completers, and during the study, finally they dropped out because of progression of disease and they had to go to surgery and they had to discontinue the drug.

So it is very important to consider these two populations, because otherwise you will miss some important information.

So the questions are -- the answer is, yes, it's obvious that we have to take into account information coming from the dropouts.

DR. JOHNSON: But, Maxime, if you have a case like yours, when you analyze the dropouts across arms and

you can see differential behavior, you know you've got the answer. I mean, you know that you've got an influence that you have to account for. In other words, your inference can't hold unless you -- I mean, the success of the drug could be due to the differential dropout behavior and not to the intervention. That's the possibility that's brought up if you see this differential behavior.

The converse doesn't work, though. If you don't see differential behavior in the dropout arms, it doesn't assure that the dropouts are non-informative and, hence, can just be ignored, which is what all the imputation techniques do.

So I don't frankly think there is an answer to this, except sensitivity analyses, that I know of.

DR. ABRAMSON: Janet?

DR. ELASHOFF: Well, it seems to me that if the dropout rate is that extensive, it has to be built in at the beginning as some kind of failure. The issue is to what extent you think most of the dropouts are really treatment failures or not versus something else, but that extensive, it has to be planned up front, and I wouldn't even try and use sort of a missing data technique. I think you'd have to describe them as treatment failures if they drop out.

DR. DOUGADOS: Two comments, the first one

relating to the comment from Kent concerning is it only a question of sensitivity. I am not sure. If we come back to the discussion we had this morning concerning the natural history of osteoarthritis, probably there are two or three different profiles, and one of these profiles could be very rapidly destructive arthritis, and from a physiological point of view, these patients have probably a disease which is different than the conventional OA, and we can easily imagine that the drug will be efficient on this particular subgroup or will be efficient on the other subgroups. So it's not only a question of statistical analysis.

DR. JOHNSON: You can analyze and use baseline joint-space narrowing as a covariate, and if you do that, if your analysis still holds, then that would increase its credibility. No?

DR. DOUGADOS: That will be not sufficient to understand the efficacy of a compound.

DR. JOHNSON: No, I don't think it will be sufficient, either. It would just increase your usefulness. What was mentioned here about having them all count as failures, we've sometimes done that, sort of the worst-case scenarios. You could even argue that a placebo dropout should be considered a success and a drug dropout be considered a failure, and if you win by that, you're

home free, but it takes an incredibly robust effect to have that.

What we did with lufludamide was a gradation of that regimen, actually. It's sort of complicated, but we tried to see how deviant the drug effect was in the dropout arms and still have the global analysis hold to .05.

DR. DOUGADOS: That is also what we have done in some of our studies, but also there's another possibility to take into account a missing value, to use the life table analysis that is more robust -- that is, to define what is a clear, relevant progression, yes or no -- and then if you conduct a 3- or 4-year study with rhetorical evaluation once a year, you can calculate the percentage of progression over time, and you define the event as the progression, and then you take into account the dropout, which is much more robust than the -- there are a lot of possibilities, but I think at least at the beginning we have to take into account the fact that the percentage of dropouts during the study will be quite high and very similar across studies, because, Paul, you conduct a 2-year study and you have 54 percent drop out, and we had 46 percent drop out. You see, it's quite similar.

And the information is important. I don't know the answer how to analyze -- that is, the sensitivity analysis, the life table analysis -- but it will be fine if

it's possible to clarify this point.

DR. JOHNSON: Ken, you had smaller dropouts, though. What's your calculated overall dropout rate going to be? Have you done that sort of extrapolation?

DR. BRANDT: We projected, I think, 25 percent over the 30 months of the study, and we differentiate dropout from loss to follow-up, because even those patients who drop because they get tired of participating or have a side effect or so, we are still bringing in for the 16-month and the 30-month radiographs. To date, we've had, I think, less than 10 percent lost to follow-up, meaning that they didn't return for the 16-month x-rays.

So there's a difference. If they've been on drug for 2 weeks and drop, that may not be so interesting. But if they've been on drug for a period of time and then discontinue drug for whatever reason, we make every effort to get the 16-month and the 30-month follow-up radiographs for analysis.

DR. ABRAMSON: Any other comments?

DR. HOCHBERG: But your study is sort of designed to have a low dropout rate, because you've got a run-in. So what percentage of subjects do you lose during the run-in phase, who either don't come back for the visits or are not compliant based on the computerized caps?

DR. BRANDT: We're kicking out about one in

four, and I think that that's very helpful. There's also a difference -- I don't want to make an assumption on what our dropout rate is analogous to the other studies, because Paul's 2-year study was placebo control against diclofenac, and it was a pain issue. Here we're talking about studies of structure-modifying drugs, where we're providing symptomatic therapy all the way along. It may not be great, but we're doing it, we're rolling it over, just as in the real world. I think there's a fair difference between those two issues.

DR. DOUGADOS: The question of the percentage of dropouts probably will decrease over time, because when we started a structure-modifying trial in the beginning of the 1990s, both the investigators and the patients did not see the differences between symptoms and structure. It was exactly the same in the field of osteoporosis, that when we conducted the first trial in the field of osteoporosis, the number of dropouts was important, because the patients needed to understand why they did not improve within a few weeks. Now it's easier to conduct a structure trial, because the physicians and the patients are aware that it's completely different than a symptomatic trial. But 8 or 9 years ago, it was not the same situation.

DR. BRANDT: I don't know if you'll recall the last slide I showed this morning, but we were very

encouraged by that, indicating that two-thirds or so of all of our dropouts have taken place in the first 6 months. We're encouraged because we perceive that as something that can be addressed by the principal investigator and the nurse coordinator in each of those centers trying to get people to hang in there, unless there's a real good reason not to.

DR. ABRAMSON: Paul?

DR. DIEPPE: I agree with everything that's been said, and the sort of 50 percent dropout rates that we got, we should be able to do very much better than that now. And agreeing with both Ken and Maxime, our experience is that most of the dropouts are relatively early on, so the curve flattens out, and I'd expect that.

But having said all of that, there's still going to be a significant number of dropouts over these long-term studies, partly because of the comorbidity issue that I raised this morning.

DR. ABRAMSON: Any other comments on point 1 here?

(No response.)

DR. ABRAMSON: Point 2, "Should a dropout exit x-ray always be mandated in the design?"

DR. DIEPPE: Yes.

DR. ABRAMSON: So moving right past the break

to assembling the evidence. Distribution of evidence from various OA sites: knee, hip, hands, and spine. Should there be one trial of knee and one trial of hip? I assume that's for the same medication, for approval. Should a drug require -- for approval for either? You're asking whether there should be --

DR. JOHNSON: We worked under the assumption that you would generalize to all OA.

DR. ABRAMSON: I see.

DR. JOHNSON: The Europeans actually want hand evidence, as I recall.

DR. ABRAMSON: Dr. Dieppe?

DR. DIEPPE: Mr. Chairman, I thought we'd done this one. I thought we'd answered this one. My view is that they've got to be done separately.

DR. ABRAMSON: I guess the corollary question is, can a drug just do one and go for indications for one anatomic location, or not?

DR. HOCHBERG: Well, I would vote for that, if I had a vote, and feel that the studies should be done in separate joints so that a compound could be registered for slowing progression of osteoarthritis of the knee, and a compound could get registered for slowing progression of osteoarthritis of the hip.

DR. ABRAMSON: Right, and not ask for a claim

for all osteoarthritis.

Dr. Dougados?

DR. DOUGADOS: Within the GREES group -- that was a group with academicians and representatives of the European agency -- we discussed this point several times, and we also discussed it within the Osteoarthritis Research Society, and the conclusion is that we are for main localization in back and even knee. It's clear that the spine should be at this stage considered completely different, at this stage, because in fact there are some in vitro and in vivo data showing that there is some relation between the disc and the cartilage. But at this stage it would be considered different.

What about hand and lower limb joints? There are some data suggesting that hand osteoarthritis, probably the natural history, the physiopathology is different, and it has been considered as different. There is the recommendation that if you want to get registration for hand osteoarthritis, you should develop on hand osteoarthritis.

The main question is related for hip and knee. So should we recommend one development for the knee, together with one development for the hip? It's obvious that everybody agrees that one clinical trial should be focused on one single joint -- that is, either knee or hip

-- but what about the labeling, what about the indications? It has been proposed to conduct either hip or knee or one hip, one knee, and to get the labeling osteoarthritis of the lower limbs. But it was a long discussion, and I don't know what is the opinion of the FDA experts.

DR. ABRAMSON: I think an argument could certainly be made, based on this morning's discussion about the heterogeneity and the comparability of groups, one could make the argument, as Dr. Hochberg was suggesting, that a drug could be developed for a target indication, such as for knee or hip, but unless they want an indication more globally for osteoarthritis, not necessarily having to meet the standard of showing both respond.

DR. WITTER: If we were to take that approach of a knee registration versus a hip registration, then if one is going for the knee, do we then make sure that there's no worsening in the hip in terms of structure?

DR. ABRAMSON: That's a good question, but I don't know.

Dr. Dieppe?

DR. DIEPPE: Well, in an ideal world, yes, but maybe this isn't that difficult, because most people either have knee disease or have hip disease. It's only a minority who have both of the lower limb sites involved. I mean, I can't give you precise figures that make any sense,

because they actually vary across the different populations that have been studied. But it's one of the arguments for treating them as different diseases, because most people either principally have knee disease or principally have hip disease. So I wouldn't get too hung up on this one.

DR. ABRAMSON: If you follow where you needed to have both hip and knee, what would happen if you showed efficacy in one joint and not in the other?

DR. WITTER: We'd come to you.

(Laughter.)

DR. ABRAMSON: All right.

Yes, Dr. Harris?

DR. HARRIS: I must say from the perspective of somebody treating a patient and not really caught up with the sort of high levels as you guys are, we have somebody with OA, and I think from the perspective of the treating physician, one might like to say it is OA of a weight-bearing joint, hip or knee, and we'd like to treat. It then becomes awfully complex if in fact it really is an indication just for a single knee. So then I guess we end up not really being able to treat all the OA of the weight-bearing joints. It becomes awfully complex.

I do grant you that the pathogenesis may be different. I do grant you that even if one decided to say, okay, we'll look at hip and knee separately, still at least

I would want that the study include hip and knee, though recognizing that one may get better and the other might not. But from a treating perspective, then we'd end up with this thing for the hip, this thing for the knee, this thing for the fingers. See what I mean?

So hopefully we could at least encourage people to do maybe a hip and a knee, even if in fact you're going to analyze them separately.

DR. ABRAMSON: Other comments?

DR. JOHNSON: Those who know the epidemiology, have there been scenarios where interventions have been able to be differentiated knee versus hip in the OA world, in the epidemiologic work that exists? Do you know of any?

DR. DOUGADOS: Let me rephrase the question.

Marc or Paul, what is the level of difference between hip and medial femoral-tibial joint versus medial femoral-tibial joint and patellar-femoral joint? In other words, within the knee, my feeling is that there is more difference between several compartments -- that is, the patellar-femoral versus medial tibial-femoral -- than between hip and tibial-femoral joint. In other words, if you want to go into the details, you will have to go into the details.

(Laughter.)

DR. DOUGADOS: And be careful. The gender will

be important, the BMI will be important.

DR. HOCHBERG: Maxime is right. There are differences in risk factor profiles between patellar-femoral joint disease and tibial-femoral joint disease. There are some differences in risk factor profiles between hip disease and knee disease. And there are differences between bilateral hip disease and unilateral hip disease, and a lot of the hip disease in the population for epidemiologic studies is not the super or lateral hip disease, which is the risk factor for progression and for total hip replacement, but it is more medial disease. So it's a real gamish.

DR. JOHNSON: But there aren't any drugs that work just on one and not the other, are there?

DR. HOCHBERG: Well, I don't think we know that. Certainly it appears that the studies of non-steroidals and other -- well, I don't want to say that. The studies of non-steroidals and most of the symptomatic therapies have combined patients with lower limb osteoarthritis into a single unit, and when we did systematic analyses, we had difficulty finding large numbers of trials which looked solely at hip OA patients or solely at knee OA patients or, in trials that combined them, where they were reported separately. So when we submitted a paper, one of the questions was, "Well, why do

you have so few trials? You must have obviously missed a lot," but in fact most of the trials combined patients together as having OA and didn't present the data separately.

But as clear as we could identify, of those agents which were studied in both types of patients, there were consistent results.

DR. JOHNSON: Consistently effective in hip and in knee.

DR. HOCHBERG: And in knee.

DR. JOHNSON: For symptoms.

DR. HOCHBERG: Yes.

DR. JOHNSON: But I take your point. I mean, I guess you're arguing that with the structure maneuver, the biomechanics might inflate some kind of small difference that we just haven't seen.

DR. ABRAMSON: Maxime, and then Paul.

DR. DOUGADOS: To my knowledge, I am aware of a single drug which acts differently at the hip than the knee. That is a placebo. The placebo is less effective on the hip osteoarthritis than the knee osteoarthritis. That is coming from the database we have within the Osteoarthritis Research Society. Otherwise, all the other drugs usually act similarly -- I'm speaking about symptoms, not structure -- at the hip or the knee, except the

placebo.

DR. ABRAMSON: Paul?

DR. DIEPPE: Well, like the others, I don't have any data that shows drug differentiation between the two sites, but, of course, physical therapies do, and that could be the key thing with structure, as you're saying yourself, because, of course, the physical measures we apply to the hip are quite different from those for the knee, because we're trying to attribute them directly to biomechanical issues around the two joints that are completely different and, of course, are structurally dependent.

So I think when we're talking about structure modification, the chances of there being differentiation is much higher than when we're talking about symptom-based changes of drugs.

DR. WITTER: Just a clarification from Maxime on your point. If I missed it, I'm sorry. Placebo response in the hip is better than the knee?

DR. DOUGADOS: Is lower. It's better in the knee than in the hip. If you look at the mean change of pain during the study with regard to the localization of the osteoarthritis, they have a better response to the placebo at the knee level than the hip level.

DR. ABRAMSON: Dr. Harris?

DR. HARRIS: Theoretically, as I understand it, in terms of companies developing agents to prevent structural deterioration, I presume that the theory is that there will be some sort of agent that prevents the sort of deterioration of cartilage, and if that indeed is the thinking by which we are going about this, then one might argue that to some degree, although there are biomechanical forces that affect the deterioration of cartilage, that indeed by looking at a drug and a placebo, hopefully with large enough populations of patients, you're going to cancel out some of the biomechanical variables, and presumably might get relatively decent answers with respect to the question as to whether cartilage, whether it be in the hip or the knee, deteriorates less with the drug that protects cartilage.

Really I don't know if I'm making myself clear, but the idea at least in terms of the development of the drug is directed certainly at the preservation of cartilage, and I guess I should ask the question, is there any way -- presumably it's the same cartilage in the hip, same cartilage in the knee, the biomechanical forces that affect deterioration are different, but if you have a drug and a placebo, then you should cancel out the biomechanical forces and end up then with just the target organ and the drug that in fact delays the progression of that target

site.

DR. ABRAMSON: I would think that would be the prediction. I guess we just don't know how much those altered biomechanics will influence the efficacy of the drug at the hip or knee, so I guess they have to be analyzed separately. Even though theoretically its mode of action would lead you to predict that it would be beneficial at each place, I guess one doesn't know that until the studies are completed.

DR. JOHNSON: A quick clarification from Maxime. The placebo response was greater in the knee, but was the drug response proportionately greater also, so that the difference was roughly the same, knee versus hip?

DR. DOUGADOS: The answer concerning NSAIDs and concerning the pain, the treatment effect was the same.

DR. ABRAMSON: Dr. Elashoff, did you have a comment? No.

Dr. Dieppe?

DR. DIEPPE: I just wanted to pick up on Nigel's point, which I think I understood, and just make a couple of comments, that although most of the drug developments that I know about at the moment are directed to the cartilage, not all of them are. Some are directed to the synovium, and some are directed to the bone. I'd like to see more of them directed to the bone myself.

But having made that cheap comment, I think the worry is that this issue that we've talked about, say, with osteophytes and stability could be completely different at the hip and at the knee. So I think there is a real chance of actually getting quite different responses at the two joint sites.

DR. ABRAMSON: Dr. Brandt?

DR. BRANDT: Also, it's not true that cartilage is cartilage is cartilage. There are some studies from 1970 by Joan Wingham and Helen Muir looking at normal hip cartilage from humans, not arthritic. In some individuals, looking simply at the surface layer, 50 percent of the dry weight was collagen; in other normal individuals, 90 percent of the dry weight was collagen. It was like an armadillo's skin, which has obvious implications for the cells underneath that. You can only say you hope that randomizes out, but that's a bit of a crap shoot. And there are a lot more differences than that. That's just one they found in those patients.

DR. ABRAMSON: I guess that begins to get at Question Number 2 in this category, as to whether one designs a study that includes both in a single large study or has separate trials. You need to be able to analyze the outcome in each area, because it may be that the drug is efficacious in one joint and not in another.

Anybody have any other comments about Question 2?

DR. DOUGADOS: Can we have first some clarification about the subset? What does it mean? Patients, demographic data, or radiological characteristics? That is, super or lateral versus medial or medial versus lateral? I don't understand the question.

DR. JOHNSON: It's just hip and knee, so you don't enroll 95 percent knee and 5 percent hip and try to draw some conclusions from 25 patients with hip disease. That's all.

DR. ABRAMSON: So basically you're asking either have separate trials on hips versus knees or have a trial where you enroll people with both and just analyze it.

Other comments on that?

DR. JOHNSON: The European document requires generalizability to all joints, right? Knee, hip, and hand. They don't acknowledge some kind of sensitivity about new disease subsets here at all, I guess, do they, in their document?

DR. DOUGADOS: I am not the European agency.

DR. JOHNSON: Oh, come on.

(Laughter.)

DR. DOUGADOS: But if I remember correctly,

hand is completely different, spine is different, but the recommendation is to conduct the trials at either the hip or the knee, but not to combine in the same trial both localization. That's a main recommendation, not to conduct a trial mixing hip and knee, either to conduct hip and another one on knee.

DR. ABRAMSON: Dr. Brandt?

DR. BRANDT: But if I remember, without going back to look, for generalizability, Kent, it required hand and a lower extremity joint.

DR. DOUGADOS: If you want the labeling OA -- that is, efficient in osteoarthritis -- you should have a separate trial on hand osteoarthritis and spine.

DR. JOHNSON: Yes. I'm sorry. They do recognize the subsets. I'm sorry. I recalled it wrong.

DR. ABRAMSON: Okay. And I think, Maxime, you really began to address Number 3 over here, which is systematic evidence for hand OA and spine OA for generalization, but not as a requirement for hip or knee.

Kent or Jim, are there other outstanding issues that you would like to -- and Marc -- cover?

DR. HOCHBERG: Well, just a follow-up to the last one. As Paul mentioned earlier that most people with knee OA don't have hip OA and most people with hip OA don't have knee OA, most people with lower limb OA have hand OA.

So the sponsor should be encouraged to measure hand OA during their studies looking at the effect of disease-modifying drugs on knee OA or hip OA.

DR. DOUGADOS: That was also the recommendation of the GREES group. That is, if you conduct a clinical trial of symptomatic knee OA, since the probability is that there will be a concomitant hand OA, you should take this opportunity to get some information concerning the structural changes at the hand level, even though the primary objective is knee osteoarthritis.

DR. ABRAMSON: Paul?

DR. DIEPPE: I agree with that. I think Marc's raised an important point. It's absolutely right. And just to make the additional point that it's actually much, much easier really to get fast readout on the hand than it is on the knee or the hip, because there are masses of joints, and addition of new joint sites, just an on/off measure of x-ray change can be quite good over a couple of years to pick up change there.

So it's an important opportunity in any of these studies, particularly with knee, because the association between hand and knee is stronger than the hip/hand association.

DR. WITTER: So is structure of the hand looking at something else, like pos-scan or something like

that, or structure you're referring to?

DR. DIEPPE: Structure.

DR. ABRAMSON: Any other comments?

(No response.)

DR. ABRAMSON: I really want to thank the panel and especially our guests, Drs. Dougados and Dieppe, who have really made great contributions today, and thank the agency and members of the audience. Thank you all very much.

(Whereupon, at 3:18 p.m., the meeting was adjourned.)