are then consistent with what we were looking at in the other measures of chronic pain.

[Slide]

So, in my conclusions, a responder analysis for pain randomized controlled trials would make sense. I would never suggest that we do it in the absence of data. I would never suggest that we prospectively put it together and then set out to validate it but that, instead, it be developed over time using perhaps a particular product and validating it from Phase II data into Phase III final randomized, controlled trials. Or, perhaps we would be able to work on it as a concerted effort with a bit of help from meta-analyses. Unfortunately, most of these domains have not actually been assessed even in recent clinical trials of pain relievers and that will limit a lot of what we can do post hoc. I think this represents minimum number of required domains. We certainly want to use validated instruments. As I have mentioned before, several different components have to be included.

As with other responder analyses, it could be required that the majority of them showed improvement but not that all would be required to

show improvement in the domains we are talking about here. As Dr. Simon had proposed, three of those five would be improved. It could be added that there should not be deterioration in the other two, or that could be omitted. The degree of improvement proposed could be based on MCID values at least for those instruments that we have.

When we know that these different domains are not closely correlated in responses, then we know that we have both a very robust clinical response when we get a responder analysis that is positive, and that we have additive statistical power which allows our sample sizes to decrease considerably. That certainly has been true in rheumatoid arthritis and, hopefully, it will be true in some of these chronic pain studies.

[Slide]

At any rate, I would just say that there is a rating scale in the "San Francisco Chronicle" for movies, and so on, which has to do with the little man and whether he is falling out of his chair or whether he is asleep. If he likes the movie he is jumping up and down, and if he hates the movie he is asleep. Perhaps some day, after we make all these evidence-based decisions, we can

develop a universal quality of life scale. Thank you very much.

DR. FIRESTEIN: Thank you very much, Vibeke. Does anybody have any specific questions about the instruments? Steve?

DR. ABRAMSON: Vibeke, a question that I guess that you have dealt with and the FDA has begun to think about, but have you lumped together diseases like RA and OA and these other pain syndromes, particularly in RA where we have mechanism-based therapies? So, if you treat with steroids or anti-TNF blockers you get a very nice response on pain. Obviously, we are going to need to sort out when we look at diseases like RA what it is that we are measuring.

I guess the related question to be grappled with is that we will have pain indications for OA that are separate from indications for the treatment of OA. I think those are two separate questions, but I guess I am mostly curious about how rheumatoid arthritis would be included in these kinds of studies.

DR. STRAND: Well, for brevity I did not include the COX-2 data in rheumatoid arthritis but, in fact, you can show very nice improvements by

ACR-20 responder analyses and also by SF-36 and HAQ even with a medication that we would consider to be largely a pain reliever.

Now, the magnitude of those improvements is not as great as we see with our DMARDs or our biologics but, in fact, most of the time patients are on background therapy with those agents. So, there is still some incremental improvement when those patients have been taken off whatever anti-inflammatory they were taking and they flared, and then they would go into these trials.

I think the other part of that is that when you see some of the improvement with the COX-2s in terms of morning stiffness, which we consider to be not a good component of responder analysis because it wasn't sensitive to change, and you see that the morning stiffness can be completely abrogated in some of these clinical trials you realize that we are again still looking at multiple dimensions of a multidimensional disease, and that the treatment of the inflammation, either by an ostensibly mild agent or even a much more significant agent, really impacts many of these domains. So, there is a lot of physical function and there is a lot of

health-related quality of life that is clearly impacted by pain. Does that get at the question you were asking?

DR. ABRAMSON: Yes, I think that is part of it. I guess the other is if a drug has an indication for OA, is it possible then to mine the data on the pain aspects of the studies that allow approval for OA and have a separate pain indication? We need to cross over what we are looking at in some of these clinical trials.

DR. STRAND: Well, I would certainly think that we could try that. I mean, I think that it has to do with the risk/benefit profile of the product as to whether you would even argue that a DMARD might be a pain reliever or might be usable just in RA but, say, OA. I think we could consider this the same type of thing and, clearly, when you look at the data in OA that I showed and the data that we just talked about in RA with the COX-2s and the data with the COX-2s in various other pain models, that is true.

The other side of it is I can't imagine that if we affect structure significantly either in OA or RA without a lot of other symptom modification that we won't ultimately still see

improvement by patient-reported measures.

DR. FIRESTEIN: One of the questions that comes up, and you addressed here to an extent, is whether these domains must not be closely correlated if they are going to be useful. This has come up again and again with regard to especially the arthritis clinical trials where the ACR-20 or even pain measurements are very closely--you are going to say no? Well, in early RA the HAQ scores do correlate reasonably well with pain. In late RA it is primarily with erosions and joint damage.

So, the issue is whether or not these are independent variables or whether they are dependent variables, and how one takes that into account when trying to set up an instrument for measuring this.

DR. STRAND: Our definition is different around close correlations. The ACR criteria, with the exception of tender and swollen joint counts, correlate with each other no better than an 0.4. In all of the x-ray trials physical function HAQ, sed rate, CRP, ACR-20 have not correlated with x-ray any better than an 0.4 and usually less. Even the tender and swollen joint counts that are considered to be obviously appropriately changing

together have a correlation of no better than around 0.7. So, I will defer to the statisticians around that, but that is one of the reasons why we have been able to decrease the sample sizes.

In terms of x-ray, we don't actually see correlations with HAQ scores until we are looking at very long disease duration, and although HAQ scores correlate very high in early disease patients, they go down very, very quickly when they get their first DMARD. So, I think we are just differing about the correlation coefficients.

DR. FARRAR: I want to address Dr. Abramson's question from the following perspective, which is that I think that one of our statistician colleagues indicated that looking at the outliers can be very informative. From that perspective, for a broken femur and intramedullary rod is a pain medicine with a very slow onset but a very long-acting action.

I think your point though is well taken in that when we are treating a disease as a primary disease we clearly affect all of the symptoms associated with that disease and, hopefully, with Clifford's help and Mitchell's and others, we will be able to look at it from a mechanistic

perspective and know whether we are treating the disease or the pain process primarily. However, I think it would be very reasonable to say that a treatment for RA that improves the disease could say in its labeling that it treats pain. However, it would not then end up meeting the criteria for treatment of a broken bone or treatment of other things where we would also want to be able to use it.

So, I think as long as we restrict and are careful about how we label what the drug is treating and, to the extent that we know, how it improves the overall symptomatology, then we won't have that problem.

Discussion of Point # 4

DR. FIRESTEIN: One of the items that we were asked to comment on is item number 4, to discuss the domains and responder indices, and address whether they adequately address the issues of efficacy or safety. I would open that up for the discussion. Obviously, Vibeke covered quite a bit of this already. Are there other comments?

DR. KATZ: Just a question. I wonder what people think the best way is to measure side effects in these trials and how important that is.

DR. FIRESTEIN: Any comments? Yes, Vibeke?

DR. STRAND: Well, we have our adverse event reporting system which I do not want to change, other than to improve it. But I think we really do need to have some type of a patient assessment, reported assessment of both the positives and negatives of whatever intervention they have undergone and they can weigh that. Perhaps we do it best with a utility measure, but I certainly see subsuming adverse events into that because then it is in the eye of the beholder or the experiencer how these adverse events truly impact and should be weighed in their therapy.

DR. FARRAR: I think there are a couple of things I would like to say about that. One is that one person's side effect is another person's effect. Just to make the point, if a drug is very sedating it may be a very good sleeping medicine and, you know, one can even look at nausea and vomiting and say for ipecac that is the effect that we are looking for.

So, the point is that the really isn't a difference in looking at side effects and effects. The measures are very often the same. I think

though that the point was just made by Dr. Strand, which is that we need to allow patients to tell us what is important to them, and that asking merely how much of this do you have, or how frequently do you have it doesn't get at the issue.

In a nice scale that was designed by Russ Portnoid to look at systems, he asks how often, how bad is it, and then how much does it bother you? This is brought out by examples of patients that I have treated for pain for whom the pain is a 10 and, yet, as soon as they develop a little bit of constipation they go off the medicine because the constipation is worse to them than the pain was. I think it is important that we give patients the opportunity to indicate whether or not they think that side effect is important to them.

At the end of the day, I would have to argue that you need to allow the patients to integrate that information. I think it was said before that we can come up with lots of models, but none of those apply to every patient. A suggestion might be the following, which is that I certainly would want patients to think about all the various pieces that go into how are you doing, like you might ask them in SF-36, and at the end of the

SF-36, so you collect all that data and you have all that for subanalysis, but at the end of the SF-36 you say considering all of the above, are you better, the same or worse than before I started the medicine? That allows the patient to integrate all of those different answers. We have assigned values to each of them; we have dictated that pain is a zero to 10 single measure in the SF-36 and that there are three measures of being able to move. So, we have said movement is three times as important as pain by the way we analyze that study. If we allow the patient simply to integrate that for us by saying overall, in terms of your pain, considering all of the above, are you better, worse or the same we are certainly gaining a sense of information that we don't get in any other way.

DR. FIRESTEIN: Isn't that essentially what a visual analog scale would provide in addition to these other instruments?

DR. FARRAR: You can ask the question any way you like, and a visual analog scale would certainly do it. From a global perspective, there is evidence that a balanced scale is better so you want to allow as many down steps as up steps to really get a balanced view. People tend to look at

the middle of a scale and then go one way or the other.

The other thing is you don't need to ask globally how are you with regards to the world. I think the issue was brought up before that your food status, your money status and your children status and all those things certainly play into it. You can ask globally is your pain better, much better, very much better or worse, a little worse or much worse and get a global response integrating the things you want.

DR. FIRESTEIN: Would that not be the gold standard for an approvable agent? If the other items were all very positive, if you were trying to assess whether something is an analgesic, isn't in the end whether their pain has improved the most important measure?

DR. FARRAR: I would agree, and I think you have stated the two important features, which is if you got the full measure of all of these subcomponents and at the end of the day you said, you know, are you better and they said I am spectacularly better but all of their others were saying they were worse, you would have to wonder about whether the questions were constructed

correctly. But as long as everything is at least consistent, I think that the gold standard is then overall are you better, worse or the same.

DR. STRAND: I would simply second that because we are looking for a robust response, therefore, we want to see it along a variety of components. It could be made so this was the primary outcome provided the others showed improvement or no deterioration.

DR. MAX: Vibeke, there is some indirect evidence from pain scores from large groups of patients in pain clinics from Jenssen and MrFarlan, in Seattle, that because of fluctuation in pain from day to day a mean of at least seven measurements over a week is more robust and may, in a clinical trial, theoretically allow half the sample size as a single measurement on the last day. But I haven't seen any such data in clinical trials. Do you want to comment on whether a single pain measurement on the last day or an average is more robust?

DR. STRAND: I will actually let Dr. Farrar comment on that in one minute because my experience is very limited with pain trials. But in terms of looking at area under the curve

analyses, for instance, in RA trials there are a lot of baseline disease activity changes over time, and that is why we typically get two pretreatment values to give us a baseline, both an over time analysis area under the curve or a landmark analysis where you are looking at responders versus non-responders at the last visit, where all-cause dropouts are considered non-responders, show very robust findings and actually reflect what we are looking at. So, I agree it could be done either way provided there is a value being given to keeping the patient in the trial.

DR. FARRAR: I think that there are sort of three ways of looking at that. Mark Jenssen has done some spectacular work looking at the robustness of different measures he looked at. I think that, clearly, if you can reduce the sample size that may be seen as being of importance. Obviously, the talk we had yesterday about how valid the measures are on a day to day basis would be important in that evaluation.

But I think the question really gets back to something that Dr. Simon said before, which is that with a sufficient number of patients you can prove anything is statistically significant. I

would raise the question of if you find that you can get a smaller difference to be statistically significant, which is really what we are talking about--when you say cut the sample size, what you mean is I can use less patients to find the difference, which is what they have shown. The argument has been made that the VAS is more sensitive than the ten-point scale. There is no question that it is; no question.

However, in studies that have been done, as you know, the variance is something like 21 mm. So, if your variance is already 21 mm, who cares if you can find a difference of 5 mm on a 100 mm scale? Because a 5 mm scale, at least in pain management, I would argue is not clinically important difference. If it was in sepsis and you are providing benefit in terms of mortality, improvement in mortality, I would argue five percent is of tremendous importance. But in terms of symptom management, I wonder whether being able to detect a 5 mm change versus a 10 mm is of any particular use.

DR. MAX: Let me respond to that. We pointed out that there is essentially no data looking in pain clinical trials chronically to

compare the sensitivity of what we are saying is the most important value, reduction in pain. The only data that I have ever seen--thank goodness for the rheumatologists--a couple of years ago Nicholas Belamy published two studies in rheumatoid and osteoarthritis where he gave people 11 different scales and he found that the most sensitive were the VAS, the zero to ten point scale, and scales that had only four points were cruder and had less power.

So, I think we are crying out for methodological studies to see if just averaging an area under the curve or taking a single last day measurement is important. John, I would agree with you that to just take a few patients could be misleading, but I think a more efficient, reliable scale is always better because you can take the same number of patients and get more subtle differences, and perhaps prove that mechanistic subsets exist. So, this is the question that I would suggest to you needs to be answered, particularly if it is our first outcome.

DR. FIRESTEIN: Dr. Anderson, and then Dr. Goldkind and then Dr. Elashoff.

DR. ANDERSON: On this issue of seven

measurements allowing you to have the sample size, I think that is likely in most cases to be an exaggeration because area under the curve analyses have been done in rheumatoid arthritis and compared with change during the trial, just looking at the beginning and the end. Although you get some improvement in power, it is not that dramatic. You know, you always want to have, of course, the most precise measure of the outcome that you can, but I wouldn't count on it to halve the sample size.

DR. GOLDKIND: I just wanted to note that the term robustness and sensitivity are different terms. I think that we have seen examples in the agency where using end of study, just a landmark analysis in chronic pain, created a p value that wasn't there--I am sorry, that an area under the curve did where a landmark did not.

The issue still remains though whether something is overly sensitive, or sensitive to irrelevant changes, or whether they are meaningful. When you are looking at how to best identify a metric that will help mechanistically, I don't think that the kind of data that we are talking about now will help in that regard. You need to see how the model or the endpoint that you are

using to assess the mechanism is affected by time. Dr. Lu's presentation yesterday I think pointed out that, in a sense, the two metrics, a landmark versus an area under the curve, give you different ways of looking at the same picture and it really depends on what you are interested in. I think one of her points was that both of them add value. In a chronic condition you want the landmark to show a difference. On the other hand, if it asymptotes out at three months and there is very little up front, it is important to know that as well.

DR. ELASHOFF: I wanted to make comments in two different areas. One is that in terms of planning your studies, it is generally better to have a more sensitive measure. The drug works as it is going to work. If you can do more studies because you can do each in a smaller sample size to demonstrate that that drug works, that is a better thing to have from an economic point of and for more science.

If you are concerned about the issue of finding statistical significance when you don't believe it is real important, then you have to address that issue in terms of clinically meaningful. It isn't an argument for using a less

sensitive measure so you won't find out what is going on.

The second point I wanted to make is that it has been stated that responder analyses don't require imputation. That is not true. If somebody quits early you still have to impute something. It is just that people are more ready to agree that you should impute the answer non-responder. It is not that no imputation is required.

DR. FIRESTEIN: Any additional comments in this area? Dr. Katz?

DR. KATZ: Just one quick comment to just again shore up what I hear as a few people's recommendation of prospectively looking at symptoms and the distress associated with the symptoms from the patient perspective. There are few papers, one written by a guy called Richard Anderson and also Marcia Testa at the Harvard School of Public Health, in Boston, looking at differences between antihypertensive therapy and another set of papers looking at differences between oral hypoglycemics. Where the efficacy of the drugs was the same, the side effects captured in a typical side effects capture way in pharmaceutically sponsored trials were equal between groups. A battery of typical

quality of life tests showed no differences between groups but a prospectively administered symptom distress inventory of something like 80 items showed significant differences between groups that then was able to predict dropouts from the trial where none of the other measures predicted dropouts.

So, there is evidence from that literature anyway that sensitive methods to detect differences in symptoms distress can actually more readily discriminate outcomes between groups than either primary efficacy AEs captured the usual way or traditionally done quality of life batteries. Maybe we should look at the same thing.

DR. STRAND: I think what we were trying to say about domains and all that, and whether it is a responder analysis or whether it is, in fact, what you are suggesting, by indicating there is not deterioration by some of these other instruments would be a very fine way of looking at the responder analysis. I think all we are trying to argue for here is that we assess multiple different aspects of the pain condition in these chronic pain studies.

DR. FARRAR: Just a very brief comment,

which is that in every academic trial that I know of we tend to prospectively collect side effect data. We ask them at every visit. We give them, you know, a 20-question scale to collect the data. In the pharmaceutical industry the adage is to basically report things that are self-reported.

I think that the concern was that in the ask mode you are going to get a lot more side effects, and that is certainly true. However, as has been demonstrated in all of the last labels that I have seen, if you display the side effect rate within your treatment group and your placebo group you can overcome that issue of having an additional number of side effects and get at this issue that Nat Katz was just remarking on, which is that it begins to help us explain why patients respond the way they do, and perhaps even get at some mechanisms that Mitchell was referring to before.

DR. FIRESTEIN: In item four it says discuss how the selection of the measurement instruments of metrics may impact the assessment of efficacy. I don't think we can specifically answer that, obviously, without knowing what the metrics are. But I think that has been adequately covered.

There are a number of additional optional points, some of which we have actually covered in some detail, including patient global issues, opioid sparing, as well as the time of onset of effect.

One of the areas that we haven't talked about, which probably we should touch on very briefly, is the placebo issue and the relative merits of active comparator versus placebo controlled studies. This is a problem that comes up frequently, and with greater frequency in rheumatoid arthritis trials where the ability to do prolonged placebo controlled trials has been markedly attenuated by the fact that we now have effective agents, and the ethics of having placebo controlled studies for longer than, say, three months now has become a significant issue.

I was wondering if we could touch on that. We talked a little bit about open-label extensions earlier, but are there any comments on the use of active comparators versus placebo controls for either acute or chronic indications?

MS. MCBRAIR: I, for one, would very much like to see reduction in placebo, or maybe not at all, especially in acute surgical pain, also with

children, and really all people. I think if we didn't have good comparators, then we would have to look at that differently, but we do. In that case, I think we shouldn't lean toward placebo unless it is absolutely necessary for some reason.

DR. FIRESTEIN: Yes, if there are rescue methods when it is clear that placebo--excluding children for obvious reasons, does that still fit--

MS. MCBRAIR: I think rescue methods certainly help but if I have waited an hour for any kind of pain medication and now I am being given something that is going to take an hour, those two hours following a surgical case, that is a long time. Two hours is a very long time.

DR. FIRESTEIN: I would agree with that, except in rheumatoid arthritis the issues are that delay of therapy can have long-term implications. Whether or not an additional hour of discomfort, and when there is appropriate consent, is a separate issue.

MS. MCBRAIR: I agree with rheumatoid arthritis. I was really leaning towards the postsurgical pain.

DR. ELASHOFF: I think the biggest issue, as a statistician, to the question of whether you

use a placebo or an active comparator is whether you are able, when you are using an active comparator, to do a superiority trial or not because as soon as you get into the non-inferiority trial issues there are some very significant statistical problems with interpreting the results of the study and it may make it very, very difficult to know what is going on, especially since the definitions of what is equivalent or not equivalent tend to be very problematic and you could easily get a situation where, from one study to another to another, you are creeping toward less and less efficacy for what you are approving. Although people worry a lot about not giving the people placebo, it is good to remember that you also are giving them something that is very likely to have fewer side effects when you give them placebo.

DR. FIRESTEIN: Go ahead, Dr. Anderson.

DR. ANDERSON: I agree with that, and I would also like to say something about post surgery trials because earlier this morning Dr. Babul, from TheraQuest presented some data from a post surgical trial which I scribbled down, I don't know if I got it all correct but it looked as though in the

placebo group--you know, it was active versus placebo, and there was a 55 percent response rate in the placebo group and 75 percent in the active group. There was more rescue medication needed in the placebo group. But I would contend that even in a post surgery trial, of course in the two to five days not the first day, there is room for placebo I think.

DR. FIRESTEIN: It is important to remember that one of the main issues we have discussed is safety, and for a compound that is in early development we don't know whether we are doing more harm than good and it may be that the placebo is the preferred arm of the study under certain circumstances, but who knows?

DR. MAX: First regarding placebo, I think analgesic experts would unanimously agree with Temple and Ellenburg's article defending the importance of placebo in early drug development. And, nowhere is it more important than in fields like analgesia. In my 20 years at NIH we have had thousands of people participate in trials and receive placebos, and they have complained about some things that have occurred during their care but I don't remember anyone complaining about

having received placebo given their chance for rescue and their consent process.

Regarding active comparators, for the reasons that Temple makes very well, comparisons of the new drug to an old drug without a placebo can be very misleading if you don't establish assay sensitivity. So, it is important in most cases to include a placebo or vary doses of one drug as well.

So far, in chronic pain studies it is remarkable that there are almost no published studies comparing within the same population drugs of two different classes. So, when we have sat down, a number of us around the table, to try and write up consensus documents on how to treat patients we have nothing to inform us. We have to go to different trials where one drug is compared to a placebo and then, in a different year and a different group of patients in a different place, another drug is compared to placebo, and because of the conditions of the study there is such a wide confidence interval that you really can't draw any conclusions.

So, I would urge the FDA to try to encourage more comparisons of a new drug to a

standard. These are hard because some people don't want to be on a standard and it may reduce enrollment. There are a lot of complex issues but it would do an awful lot for prescribing practice to have that information.

DR. WOOD: I agree with that. I think it is very important that as far as we possibly can ethically we include placebo. Bob and Susan in their article very eloquently point out that everything that we know about placebo-controlled trials has stood on its head almost statistically when we try to use active comparators. More carelessness in the trial, all the kinds of things that normally discipline us are overturned. So, I think we use active comparators at our peril in particular in an area like this. So, I think we should certainly be using placebo as much as we can with appropriate ethical and safety issues, like using escapements and so on.

DR. FIRESTEIN: Yes, Dr. Borenstein?

DR. BORENSTEIN: I just want to point out that the difficulty we have is that placebo works so well, and if it didn't work so well life would be much easier for us. The difficulty is placebo, as pointed out, is not necessarily a bad choice,

unfortunately. When that happens we have to just wonder what is happening in those individuals. So, I have no trouble when asking patients to be in my trials. It may not be the largest group but I do think placebo is something that should be in these trials, and people are willing to participate in those circumstances.

DR. FARRAR: We aren't here to discuss the pros and cons of the placebo effect, which obviously could take a whole day in and of itself. However, just a comment which is that every person every day of their lives uses the "placebo effect" to affect how they feel about what they are doing and whether they go to work because they bumped their leg or not. So, I think that the issue of whether it exists or not and what it means is important to take into consideration. As was just commented, it can work really well in certain kinds of syndromes, not so well in other ones. And, I think that the primary issue is what Mitchell was saying and what Dr. Wood was saying in terms of the need to have a comparison against something that is the least active, and that would be placebo with the appropriate controls. It is rare that you cannot come up with an ethical way to do it. Even

in a postop trial, if you are giving somebody a pain medication that is supposed to work and you give half of them a placebo, at the time of the maximum pharmacologic dose you ask them is this enough, and if it is not you give them a rescue. Most patients, as I think was said, are willing to participate in a study where they may have to put up with some pain for a period of an hour or maybe a little bit longer.

I think the second thing to mention is that I have heard today or yesterday perhaps a couple of times when people said placebo corrected trials. I don't know what a placebo correction is because the placebo effect is for free. You get the placebo effect. When you give an active drug you get the placebo effect. What we are really looking at, and the advantage of a responder analysis, is whether people reach a level where they are satisfied with the relief in pain, or whatever, and it doesn't matter what the response rate is in the placebo group in terms of trying to ascertain whether or not people are better. Right? The question is better or not better. What then matters is to decide whether the difference in the response in the placebo group is sufficiently

different than the response in the active treatment group. The one place where the placebo effect can be problematic is if you have a population where you end up at the top of a scale. If you end up with the placebo effect working in 90 percent of your population, then you are going to have a lot of difficulty showing that last 10 percent where you got a clinically important difference.

So, I think there are some issues but it is not really related to subtracting out the placebo effect. I think that doesn't get us anywhere.

DR. FIRESTEIN: At this point, Lee, would you like to summarize? Good luck!

Summary

DR. SIMON: Thank you, Gary and thanks again to all the members of the committee for such interesting discussions over the last day and a half. I actually come up here with some humility, being able to actually attempt to summarize what we talked about and I hope that you will find it useful.

There are a couple of statements that have been made throughout from people on the committee that I would like to be clear about. You know, we

are very open and we would like to believe that this kind of meeting reflects how open the division is to discuss with the sponsors and other interested parties the way drugs are developed. So, I think that is the first thing that needs to be said, and can't be said enough.

[Slide]

We reviewed chronic and acute pain, and we reviewed the concepts of the clinical approaches and the concepts of the mechanistic approaches, recognizing, of course, that the mechanistic approaches are rather nascent in development. We are not yet there and we still have to grapple with those drugs that are presently in front of us and to be soon in front of us, and have clear messages about how these drugs can be approved for their various different indications. Although we would like to believe that the mechanistic approaches are just around the corner, they are not yet there and I don't think any of the protocols, drugs and designs that we have in front of us right now are actually dealing with mechanistic issues.

[Slide]

I think this sign really summarizes what I mean by being clear. I don't want anybody to feel

like our division is giving you mixed messages. I really would like you to believe that we are giving you the real arrow to the right when it really needs to be to the right.

[Slide]

So, we discussed temporal descriptions of acute versus chronic for example, or intensity differences such as mild, moderate to severe, and we decided I think that they weren't enough to really inform us about where we wanted to go. Some of that is because of the issue of is chronic as broad as it should be, or is it too broad, and those kinds of issues.

So, we clearly need further clinical trials to define mechanisms because we can handle mechanisms better, but that is for the future, and it is unknown whether there can be a global analgesic right now for we know there are quite different mechanisms driving the sensation of pain.

[Slide]

There is clear concern that we need, as an agency, to design claims and consider proposed trial designs fostering new development, new drug development for pain. I actually think that is very true. For the chronic pain proposal, I heard

some people thought it had merit. That was again, just to remind everybody in case you have forgotten, three models, three co-primary outcomes of pain function and patient global, and it would be replicated in nature with disparate etiopathogenesis mechanisms or disease states. They were replicated, necessary, when you were doing studies in models with simpler mechanisms or not. We weren't sure whether or not it was going to need to be replicated in that particular circumstance.

And, it seemed that in the vote we took, although there was no vote but consensus building that we took, although I am happy to say I understand the camps, I am not entirely sure we got consensus. Most people said yes to pain as a measure; yes to patient global and that is a measure of clinical relevance of the response; and there was a qualified yes to function. We would need to take that into consideration of the model or mechanism or disease state that we were talking about. Obviously, cancer function or a patient with cancer who is functioning, that would be a different issue than some other diseases.

There was debate of how many different

models are required to get any type of specific claim for chronic pain. Are three different models required? Dr. Verburg suggested four models of one trial in each. Maybe Dr. Firestein resonated with that a little bit. We were suggesting three models with two replicate trials. Dr. Farrar suggested two neuropathic models and two somatic pain models. So, clearly, we will be taking back this information to think more about what we should do.

[Slide]

In that context, the lumping and splitting context is very important. We had thought we were doing both lumping and splitting because we gave the opportunity to split or lump. Dr. Abramson kind of resonated with the rigor that would be associated with that kind of approval, and it really raised issues about whether it would be iterative. You would get one indication and then perhaps a much broader organ-based indication, and then perhaps a whole disease indication, fully recognizing, however, that the daunting nature of the full, whole thing, the whole kit and caboodle may be just too much and, in fact, companies would opt for something easier, perhaps cheaper, and then off-label use would drive that and that would not

be an ideal situation. I think it is really critical for us to remember that we were providing in our proposal that opportunity, for better or for worse.

[Slide]

We also recognized and heard clearly that acute pain is not similar to thinking about the drugs that would be used to treat it. Thus, actually we are thinking about short-term analgesics rather than drugs for acute pain. The same thing in obverse is true for chronic pain. We are really thinking about drugs to be used for a long period of time and that has issues regarding safety and durability of response in trial design.

[Slide]

We learned something that I think we have consensus on, that chronic low back pain, if handled correctly, might be an indication to go for independently, or actually may be part and parcel of a much larger package. Although heterogeneous, it consists of many different processes but they can be delineated, and we could select a specific patient population with some similarity in the natural history, perhaps ignoring or removing those patients with reticulopathy or neuropathy, and

perhaps we would have a model that we could use or pain disease state that we could use for a clear indication, as well as performance of a broader label. It seemed that there was good consensus about that if we made sure that we subtracted out patients with neuropathic disease and systemic disease.

I think we heard clearly that there are two really broad patient populations that we have not dealt with very well. One is the elderly and one is the pediatric population, and we have to recognize that the elderly are quite unique. Polypharmacy is a significant issue with them. Safety issues are particularly important, and some of the elderly who are suffering chronic pain are in unusual care-giving environments. Perhaps as the baby-boomer population gets older it will be a usual care-giving environment, but we have to learn how to use nursing homes for actual study designs and carrying out studies in those areas as the patient population in them grows larger.

[Slide]

The issue of flair design was debated. Some of us had problems with flair design. It actually has been tried and true but, on the other

hand, it preselects those patients who both tolerate the drug as well as respond. A priori they have been on the drug for a period of time so there are issues about that particular problem.

We heard about possible ways to do a run-in phase and withdrawal studies, both of which have problems. The run-in phase really doesn't do anything differently than does the flare design. It suggests that you are only taking patients who are having a response and getting rid of all those patients who can't tolerate the drug. So, you have a true bias in the evaluation.

The other concept of the withdrawal phase which Dr. Laska asked me to comment on was, in fact, some concern about are the patients who get withdrawn unblinded or not based on the symptoms that emerge? So, that is an issue that I think we are going to have to think about.

[Slide]

Many of us talked about the issue of opioid sparing, although it is not dissimilar from glucocorticoid sparing, and how important it is for the assessment of outcome. It might be a good response to measure. Would it be a primary measure? Probably not. It might be a useful

secondary measure but we would have to debate that, demonstrating that the study drug works and decreases the need for opioids and, presumably, the study drug in the circumstance would have less side effects than the opioids so there would be a warranted reason for the study. The problem, of course, is that the study drug might enhance the effects of the concomitant opioid therapy, thus decreasing the use of opioids or, alternatively, decreasing use of the opioids may be due to the emergence of increasing toxic effects.

What I am constantly daunted by, and I am not really that far off in glucocorticoid sparing either, is that I don't know what it means to be sparing because I don't know if 3 mg is better or 30 mg is really sparing, and I think we have to debate what that really means. As mentioned by Dr. Wood, there is the issue of the PK change and what that would imply to the whole process.

[Slide]

We then moved on to the ABCs of acute pain, and there seemed to be--perhaps you could show me with smiles on your faces--less debate about this. This seemed to be something that you all bought into faster for good things.

[Slide]

Clearly, we want to improve the information in the label by turning from inferences evidence by PK modeling to data derived from clinical trials. That would be the multi-dose assessments. That was informed by the B of the ABCs.

We want to improve safety analysis of short-term use by analyzing long-term exposures even for drugs approved only for short-term use. There seemed to be some confusion as to whether or not, if we were going to require some chronic exposure, and maybe even efficacy trials, that that actually might mean two replicate trials or three co-primary outcomes, maybe even three different disease states. That is not really what we were suggesting. It probably would be just one trial, perhaps even just very robust and perhaps just one outcome measure but we would have to debate that and talk about it in an open fashion to determine exactly what we would want. But this was then informed by proposal C of the ABCs.

[Slide]

We clearly heard that generalizing to postop pain and efficacy from a dysmenorrhea trial

or dental pain trials really was a problem and we have been very uncomfortable with that. So, we needed to think about requiring or suggesting that not only does one do an outpatient trial in such a circumstance, but one might want to choose an inpatient model which would give a broader aspect of pain relief, thus, a bunionectomy model as well as a dental pain model.

Additional info regarding the dosing interval was needed, and that was clearly defined by B of the ABCs; more optimizing of the dosing schedule in responder versus non-responder inclusion, which I actually found to be a fascinating discussion.

[Slide]

Dose creep was brought up, and I think that it is very important. and it came up several times from the committee that we need to construct our clinical trials in a real-world way to ensure that we understand how the drugs are going to be used in the real world, and that doesn't imply open-label analysis; that just implies different ways of thinking about trial design than we have done before. Issues of longer time of use requires the chronic studies, as we talked about.

[Slide]

The discussion went on after the presentations regarding the matrix of clinical trials. Again, I think everybody around the table believed that they should inform us about real-world use and should be labeled as such.

Time to rescue should include the non-responders and that implies an intention-to-treat analysis, not just a responder analysis.

New designs with preemptive anesthesia raises the question of whether or not we should be thinking about that differently than acute pain, and maybe that is a whole other world of trial design, and all the consultants out there can start to think about that and create new business for yourselves, which is a good thing. Improved GDP and all of that.

Short-term studies, pain relief, patient global in terms of level of response for how long and when is the onset; when it separates from placebo; drugs not with onset within an hour but a very good analgesic, do they inform about some acute use? In fact, that came up several times, this idea that there is the acute; there is the

chronic; but what about kind of the middle ground? We need to start to think about this subacute use and what that really means.

[Slide]

Also, going through dose descriptions and minimum time to the next dose is informed by the time to onset. It needs also to be limited by total dose and dose ranges may be better described by quartiles of response. I really like that idea. I think that really gives us a much better handle on what this all means.

[Slide]

Lastly, but not leastly, we heard about a tiered responder analysis, informing patients and clinicians much more so than present analyses do for pain. One could see that in acute pain you could define a level of pain relief, along with the duration of pain effect within the same construct of explanation or description. And, in chronic pain it would develop an information database including efficacy, kind of encompassing pain and suffering relief; durability of response; time to retreatment or time to treatment failure; as well as function and HRQOL measures; and then also safety. So, this would be a remarkably robust data

set to inform patients about what really is going to go on with the therapy.

[Slide]

I want to close with this, and I don't really show this entirely in jest--entirely. This was actually a real traffic sign in England where they actually advertised and demonstrated the directions to the secret nuclear bunker. We don't really hold any secrets in the agency. People have come over to me and said, well, would you really talk to us? Or, can we come talk to you? Or, we have our stuff already in and we are talking about changing, are we going to be held to a different standard when we have already done all of our trials?

Well, one, you need to talk to us. Make an appointment and come in for a meeting. Call your project manager and see what the status is. I would prefer not to hear any complaints that we are not willing to talk to you. I am being very public about this. We are willing to talk to you. There are no secrets here.

Number two, we are willing to debate with you as to what might be happening in this particular turbulent time of change because, in

fact, we are trying to do, and I think you all are too, what is best for patients and to derive the most information in the most open way. So, I invite you to give us a call. Those of you that have not been in for a while and have been busy developing drugs, I really urge you to take advantage of all the opportunities to have guidance discussions because, in fact, it is much better to come in and talk to us before you come in for your pre-NDA meeting and be surprised.

So, in that context, let me suggest that we show you the way to our secret nuclear bunker and give you all the directions up front, and I think everybody will be happy.

So, thank you again very much for coming. Thank you to the committee for working so hard in helping us and informing us about your ideas. I don't know what will happen next but we will certainly have another meeting about it.

DR. FIRESTEIN: Thank you very much. The meeting is closed.

[Whereupon, at 2:30 p.m., the proceedings were adjourned.]

- - -