along the lines of the minimum time a patient should wait before taking a second dose is two hours, and that would be dictated more by the onset of action rather than the time at which the medication would run out, and that the maximum number of pills allowed in the first 24 hours is such-and-such, and allow physicians essentially to give patients the right to take enough medicine to achieve the relief that they are entitled to get in a safe circumstance.

DR. FIRESTEIN: Larry?

DR. GOLDKIND: Particularly for an opioid that may be a good model. The problem is if you have a non-opioid, there is a whole different mechanism where the dose response curve is not quite as clean. If you tell somebody, based on safety, you can take another dose in a couple of hours, we don't really know that that second dose will benefit other than the placebo effect.

DR. FARRAR: Could I respond to that? I agree with that, in which case I think the issues that were brought up before about the 25 percent non-response, or the time point at which 75 percent of the patients still have an effect would be a reasonable dose interval where 25 percent had

started to take an additional dose, as long as that is a safe dosing regime.

DR. GOLDKIND: We do get data submitted that has it in quartiles and the median is simply the one that is highlighted. It doesn't really help in decision-making. It may help in terms of approvability. It may help in labeling to have that data displayed so people know when the median will rescue. We would have to deal with the variability of whether, again, it is responders or whether it is all patients. Frankly, in the model are we going to apply the dental pain or the surgical setting to that description? We could end up with a ten-page label if we were as informative as we may discuss here.

DR. FIRESTEIN: Dr. Borenstein?

DR. BORENSTEIN: To follow-up on that point, I think part of the responder aspect may be the half-life of the drug. While in the label it may be a certain half-life, human biology, when it comes to the clinic, seems to have a much wider range. So, there are some people who say, yes, I can take this drug and it truly is once a day, and other people really say it is twice a day and I need to take it because I really experience the

lack of efficacy. So, it will have an effect partly on your response, but also if you can get the data which shows the range of what it may be in a variety of patients so you actually can tell that. Tthat actually may make for a better label, that it is a range and that when you have that you have individuals maybe on the short side and the long side. So, you may find with your dosing that, in fact, what may be once a day in some patients may actually end up being twice a day and to get efficacy for those individuals you will need to dose it that way and the drugs will have a wider range of effect.

DR. DIONNE: I was going to endorse the proposal that Jim Witter made about acute pain responders as an alternative to doing mean or median responses. We are probably at the point now where we are going to have a better potential for understanding the basis for individual variation due to genetic factors. If we have the data that we are using to analyze the range of responses, we could possibly better interpret what is going on not only on an individual basis due to the genetic variation, but also we would eventually be able to form, I think, more reasonable judgments about the

safety or efficacy of a drug.

If there was a drug that had a very effective median dose, nice duration but one out of a thousand patients had a very serious adverse response, we might be much less willing to see that as a drug for acute pain use or eventually consider it for over-the-counter use versus having the perception that this drug has significant liabilities or significant variabilities that affect its clinical use. So, if we had a formal way of doing responder analysis we could get at that variability.

The only problem is I would hope that we would derive that due to some data-driven process rather than just some sort of an opinion-driven process. It might take a couple of years for that to evolve.

DR. FIRESTEIN: You mean actually use evidence-based medicine?

DR. DIONNE: Something like that.

[Laughter]

DR. FIRESTEIN: Dr. Wood?

DR. WOOD: It is important to recognize that the duration of effect is not a simple relationship to the pharmacokinetic half-life. The

duration of effect would depend on the time for which the plasma concentration is above the minimum effect of concentration. At a high dose that might be very long and at a low dose that might be very short, both of which might not be obviously related to the half-life. So, the pharmacokinetic half-life is not a good measure of the effect and duration, and probably should be ignored, except in the sense that, obviously, a drug with a very short half-life will likely last less time than a drug with a very long half-life unless the drug with the very short half-life can be given at doses that are way above the minimum effect of concentration.

DR. FIRESTEIN: Let's spend the last couple of minutes talking about point three, which is how does one determine if a difference makes a difference. Would you like to get us going since you are the one who generated that pithy quote?

DR. KATZ: Sure. I think it is actually Yogi Berra or somebody like that. But I think it is an empiric question and just needs to be explored empirically in the context of whatever model one is looking at. John Farrar has done some very nice work in looking at clinically important difference in neuropathic pain and I think, John,

you found that it was about 30 percent reduction in pain.

We have done some work in a chronic back pain study that Dr. Borenstein participated in. In the analyses that we have been doing it looked more like 50 percent pain relief was associated with global measures and other signs that were the marker for meaningful pain relief. So, I think it depends on the individual model and it is an empiric question.

DR. FIRESTEIN: Vibeke, in the arthritis studies with visual analog scales, what have you found to be something that is significant?

DR. STRAND: I will show you this during my talk, but basically we found that it is about 30 percent, 30-36 percent, looking at correlations with patient global assessments for various other parameters, such as HAQ, disability index and so on. It is about 18 percent above placebo. As we just talked about, Dr. Farrar's work across ten trials, randomized, controlled trials in multiple different kinds of pain was very consistent. It was approximately 30 percent. By VAS, we think that the test/retest variability, if you are using 100 mm scale, is about 20. So, when you get to

about 30 you have a minimum clinically important difference. That seems to work no matter what kind of a VAS scale you are using. Again, I will show you some of that data later.

DR. FIRESTEIN: Dr. Sherrer?

DR. SHERRER: I might be assessing a rescue medication use because I think that is the patient's indirect way of telling us what is adequate if the pain medicine is adequate by itself and they don't have to be rescued. If they have to be rescued, no matter what the pain relief was, to me, it was not adequate. It doesn't mean that that drug is not useful. It may be useful in combination but, to me, if the patient has to be rescued they are telling us whatever it did, it didn't do enough.

DR. DIONNE: I was just going to add to the discussion of what is the minimally effective increment of pain improvement. We did a study in the oral surgery model with about 125 patients starting with either moderate or severe pain. We slowly titrated a nonsteroidal anti-inflammatory drug IV until they reached a point where they pressed the stopwatch, and then we had them fill out their category in VAS scales. It was startling

that it came out to be about 50 percent pain relief across the different types of pain intensity in different scales.

DR. MAX: I have two concerns about setting a minimally significant clinical difference. One is that I am afraid of approval creep. Now it is enough, given a reasonable safety record and a sense of clinical usefulness, if you just beat placebo within an acceptable alpha level. I am afraid if you establish that you need to have really 15 percent pain relief, the requirement may creep into being that the studies need to be statistically significant above that level.

Alternatively, I want to point out that it really depends upon the context and the side effects. If you had an analgesic that looked safe and had no, say, cognitive side effects, you could add it to most of the analgesics that are sedative, and even if you only got five percent or ten percent additional relief, it is cheap enough and it would be a very welcome addition. So, I would want to leave this to the case by case judgment of the agency.

DR. STRAND: Could I just clarify for a minute? I don't think we are talking about MCID

based on one outcome measurement as defining clinical response. That is why I would like to put this off until this afternoon when I present.

But I think what we are really trying to talk about is where do we see minimum clinically important differences in various parameters. The way they become useful is if you now combine those parameters that are not closely related into some type of an analysis for responder. All of this has to be done as evidence based.

DR. MAX: Yes, and it just depends comparing to the safety profile of the clinical context.

DR. FIRESTEIN: Dr. Cush and then Dr. Elashoff, and then we will take our break so that we don't have break creep as well.

[Laughter]

DR. CUSH: I just want to go back to Yvonne's suggestion, and I agree that the use of rescue medication is certainly an important measure and I think one that is useful for analysis, but I am also bothered in doing clinical trials where we use rescue medicine, especially in osteoarthritis, by the number of patients who refuse to use rescue medication despite their pain. I can't quite

explain that. I know they have pain but they continue to not want to use the analgesic medicine we give them. So, I somehow fear that we may be missing an important outcome if we rely too heavily on that one measure. That needs to be included but I don't know that it can be a primary outcome measure.

DR. ELASHOFF: Any time one feels one needs multiple measurements in order to understand what is going on, you are either left with trying to sort of put them together after the fact, after they have all been measured, or defining some arbitrary combination of them. There is always an arbitrary character to that, and if you define things ahead of time then you are liable to lose information later on. But there is always a tradeoff. There is no way to totally win this situation.

Dr. Cush's remarks about the rescue medication issue are certainly important ones. The advantage of that particular type of outcome--or at least if we don't think of it so much as rescue but amount that they would actually take if left on their own, the advantage of that kind of outcome measure is that it is directly related to the

safety issue in a much clearer way than some of the other outcome measures one might be talking about.

DR. FIRESTEIN: Dr. Simon?

DR. SIMON: Just before the break, if you will give me a minute, there are a couple of questions that arose in the previous discussion that weren't really answered by us. One was Dr. Katona's question about were there other alternative designs besides a placebo-controlled trial. That would be appropriate and, yes, obviously an active comparator would be an acceptable way to go for an acute pain trial in children, elderly, in any number of different ways to do that, background therapy, withdrawal therapy as has been done in children before, though I am not that enthusiastic about withdrawal therapy in adults despite what came up yesterday and I am sure we will discuss that part again.

Number two, there was an interesting discussion about acute pain, time to onset of acute pain, differentiation from placebo and preemptive anesthesia. I would like to point out that we are willing to consider that as an entirely disassociated issue, meaning, we have to create a label that patients understand how to use drugs.

We believe the time to onset of an hour may be important to patients as opposed to two hours, although I do not want to get into a discussion, as we did in '98 on fast, faster or fastest because, in fact, that is not really informing us anything. The reality is that there may be the need for an entirely different indication of preemptive anesthesia rather than acute pain because, in fact, that is a different issue and it would affect different patients. There are not a lot of patients walking around with a toothache who need preemptive anesthesia as opposed to acute pain relief.

The third issue is the issue of effect size that Dr. Elashoff referred to before. It refers back to what Dr. Max was talking about, which is that we have to be familiar with MCID because if we don't consider that the sponsors, not because they are bad people but because they have accrued a lot of patients in a trial, can then have enough patients to show a statistically significant difference from placebo yet, in fact, the effect size is entirely unimportant.

Part of that is bias and a take on how big is the effect size. It might be nice to know that

an effect size is evidence-based and defined by what is minimally clinically important, and that may be very important because of the number of patients you could recruit. You can't just make your study be positive.

DR. FIRESTEIN: Thank you very much for clarifying those issues, and we will take a break now. We will start again in exactly 15 minutes, 10:45.

[Brief recess]

DR. FIRESTEIN: Can the members of the committee please rejoin us? In this session we have an open public hearing. Then, we are also going to try to clarify or revisit some of the questions that were raised yesterday with regard to chronic pain indications. We have two speakers, Dr. Eugene Laska who has been allocated ten minutes, and then Dr. Nijab Babul who has been allocated five minutes, and I would like to welcome them. Dr. Laska?

Open Public Hearing

DR. LASKA: Thank you.

[Slide]

This little presentation is sponsored by Merck, whose folks I would like to thank for their

stimulating comments and stimulating discussions which led to the clarification of several issues among the contributors, their ideas, particularly Al Sunshine whose name I want to mention. The ideas here are ones I have talked about before. I apologize for repeating some of them. Lee Simon and Jim Witter and Ray Dionne also deserve special recognition because they are clearly attempting to open up the box and make the business of registration more transparent. Some day a drug company will know whether they are going to get approved before they make a submission rather than wait for the surprise of the letter.

As I mentioned yesterday, the goals of a randomized, controlled trial are to allow causal inference; to allow the conclusion that the drug is the reason for the effect we observe.

I want to add to that that another major reason for doing clinical trials is to get point estimates of very important parameters which characterize what the drug is all about. It is instructive in trying to design clinical trials to contemplate how one would use the information that comes out of them; what kind of information one really wants.

If one thinks about onset, duration and dosing intervals as if you knew the entire story, you know, the probability distribution of onset and duration and response rates, you would see that it is a complicated, multidimensional space that would be very hard to characterize. And, what we are looking to do in these clinical trials is to find very, very minimally informative point estimates which describe to some degree the amount of the effect that we are talking about, median time to onset and the like.

Too many measures, as Janet says, are not necessarily useful, and for these trials for the longest period of time we have collected data on both relief, which refers to original time, and current intensity. I am pleased to see the agency moving to the notion of dropping redundancy at least in the notion that it may be redundant in the beginning but certainly long term. Good thinking.

The same thing is true about all of these parameters. They are functions of pain intensity levels. So, again, the hyper space in which these characteristics are described is very, very high dimensional.

[Slide]

Let me start by talking about stopwatch and measure onset. I believe that it is important to eliminate the two stopwatch theme that has been used by many companies in the recent past and return to the one stopwatch approach that measures meaningful relief because I believe that that is the most useful concept that can be measured, and that the redundancy in having a second watch to try to capture perceptible relief merely adds complexity and does not really bring in enough new information to warrant or justify its use. And, I think that second stopwatch is a very useful tool, which I will mention in a second, that cay be used to look at duration.

[Slide]

Once one collects the data, I think it is important to conceptualize the ideas associated with onset as representing two subpopulations, one, people who will not respond or who have not responded; and the second, the group that has responded. That is characterized statistically by the top equation. It is called the cure model. We won't talk about it today but it has been described in the reference in the bottom of the slides. That particular model conceptualizes the outcomes as

falling into two groups, the responders group and the non-responders group.

I believe that the regulatory indications of collecting data the way I have described and breaking up the population into these two subgroups flows very naturally. The clinical trial's objective will be to estimate the proportion of patients who respond, who get this meaningful pain relief, and look at the survival distribution including the median time to obtaining meaningful relief.

[Slide]

The regulatory implications that flow from that I believe fall in two camps. One is a comparative camp and the other is a numerical estimate camp which has to do with characterizing the drug independent of another drug or placebo.

So, the first requirement would be that Pd is bigger than Pp for the placebo group. The proportion or response must be demonstrated to be statistically superior on the drug than the proportion who respond on placebo. Perhaps a minimal difference in the proportions is called for so that sample size doesn't dominate the decision as to whether there is a proportion.

[Slide]

But then the issue of whether or not a drug works within an hour or more generally within T units is characterized by the second requirement which only talks about absolutes, not comparators. That is, the median time to onset among the responders on this drug ought to be within some period of time, perhaps an hour, perhaps an hour and a half but more generically T. T, of course, may depend on the pain intensity, the model setting and a variety of other things relating to the individual and the biological response that that individual represents.

[Slide]

Perhaps more difficult to contemplate is the question of duration.

[Slide]

Let me suggest to you that the FDA's concerns about using the various interferences that are introduced by the nurse or whoever is collecting the data or deciding whether or not to give that second dose is mitigated by putting that second stopwatch that used to be used for something else, so they are around and there is no extra expense--that second stopwatch can be used to

answer the question when is the patient no longer getting pain relief.

The agency used to worry about what they called back then the minute wars of the first interview for onset at 15 minutes, demonstrating efficacy, would provoke another drug company to collect its first interview data at 14 minutes so that they could claim faster onset. Well, the stopwatch eliminates that problem and it does so here as well. It removes the bias, the interpersonal possible interference that the nurse observer or the person who could give the next medication introduces.

The estimating functions that would derive from collecting data of that sort are exactly analogous to what we would obtain in the onset story. We would estimate the survival distribution of time to rescue and the proportion who respond. Very importantly, they do not impute a value for those people who never got onset.

The question of how long a drug works after it has worked is not informed by the percentage of people or the time at which those people rescue if they never got onset. it is a different question. The answer to the question of

when shall I remedicate when a person is not doing well on the drug I gave him is a very different question from the one that asks when do I remedicate after there has been a long period of time where the patient has responded.

A number of the things that can be reported along the way are the proportion who respond at the various times that are convenient, like 6, 12 and 24 hours; median time to rescue among responders who do rescue.

Let me focus on that for a minute. It is useful to say ten percent of the patients respond, and among the ones who do--sorry, median time to rescue. Among the people who rescue, how long does it take before they need rescue? That is going to depend on severity and the like, but that informs the notion of the time to rescue and is a complement to the proportion who don't rescue. Those different arms are the reason I described in the beginning the hyper dimensionality of the outcome space when you do a clinical trial of this kind. To mix them up is to blur and lose information about what is actually transpiring.

[Slide]

The regulatory implications of choosing a

dosing interval on this basis has to do with, in my view, a compromise between the wide range of dosing intervals that are absolutely necessary, that all of the clinicians on this panel discussed in the last hour but, nonetheless, if the agency chooses to characterize with one number, I think that number is the median despite the comment that I don't want the other half of my patients to do poorly because the dosing interval is honored in the breach. So, if this is the one number you want to produce, I think you are stuck with the median and, therefore, the dosing interval is some number less than or equal to the median time to rescue.

I believe the limitation that you place on providing information in the label is a very artificial one, and the notion of posting information on the web doesn't need to be defended. You don't need to hide behind the label to describe what happened in the trials; put them out some other way. Once they are out, clinicians will find a way to use them if they care to find out the information.

So, the regulatory implications are that the percentage of patients, the second point, who need rescue is significantly less than the

proportion of patients who need rescue on placebo among the people who responded to placebo. That would need to be demonstrated statistically.

The first point, the comparative one, the absolute is that the proportion of responders is less than some fixed time point, and that is less than a half.

[Slide]

Just one comment quickly on Larry's feeling that return to baseline is a flawed metric. I think one can conceptualize this whole idea as the complement, the counterpoint to the responders analysis. If you like, this is the failures analysis and patients will return to baseline individually. The argument that the mean does not return to baseline doesn't mitigate against the use of return to baseline or no longer getting meaningful relief on an individual basis, and it is the counts of how many of those people there are as well as the time to the event that makes the game playable.

[Slide]

So, clearly informed by PK and informed by the experience of the clinical trials in the acute phase, one has to look at multiple days and the

question is what to do in that context and I had to think about it. My view is that this is not the place to be exploring dose response. In the very mild pain circumstances where pain is almost gone the next day, it makes no sense to me statistically as a statistician to impute data from day one to day two to show artificial differences which are not real.

I believe that you can only sustain the notion of what the effective dosing interval that has been proposed and see if it makes patients "happy." So, at the end of day in these mild cases there should be no need to demonstrate superiority to placebo, but the proportion of patients who require rescue ought to be smaller than some absolute number that is credibly determined on a judgment basis.

[Slide]

For more serious pain or perhaps severe pain models were PRN narcotic is required, I see no alternative to the idea of using the dose sparing property of the drug.

[Slide]

There is an old rule that every animal pharmacologist will ascribe to, I am sure, that

says if you fix dose, study outcome. If you fix outcome, study doses. In the dosing sparing setting where you use PRN narcotics you are fixing an outcome. Patients titrate to adequate relief. The only thing to study is the amount of narcotic that is spared. It is sensible and there are caveats raised by others in the group here about interaction, about promoting side effects. Remember, this drug has been studied in the acute setting. It is known to be an analgesic. Now the question is what does it do on day one, two or three and that kind of sparing relationship, in face of the knowledge from the earlier trials, is pretty clearly evidence if you believe in the hidden assumption--as Jim pointed out, there is a hidden assumption and in this case it is that there is a dose response to the narcotic being used. So, dose sparing makes sense to me as the way to sustain that data.

[Slide]

One last situation then, we are in long-term use, and I am anxious to hear the objection. If chronic pain situations where patients on placebo drop out at very high rates, once again we are into the game of projecting

forward; we are making up data--statisticians call that imputation, to justify whether the drug still works at week W where W is a big number like 12.

I think that makes no sense. It is a circumstance, again, where we are only trying to sustain the notion that this drug continues to work after 12 weeks. We are not trying to prove effective here; it is does the drug still work? The best way to answer that question is not with respect to placebo patients who drop out earlier; it is with respect to patients in whom the drug is working, it is withdrawn and superiority to placebo in a randomized, controlled trial is demonstrated.

I believe that this kind of an approach is a rational way of looking at onset and duration and choosing dosing interval. And, I thank you for listening.

DR. FIRESTEIN: Thank you. The next talk will be from Dr. Babul, from TheraQuest.

DR. BABUL: Good morning.

[Slide]

I would like to address the committee and the division on the issue of multi-dose analgesic development. This is one of the questions that the division has asked the committee to consider in

terms of evaluating analgesics in acute pain.

[Slide]

I have previously provided a conflict of interest statement and that stays on record so I won't repeat it here.

[Slide]

This slide shows the essential approach that we have been taking for the last two decades to evaluation and approval of analgesics in acute pain. Certainly from an efficacy perspective, we do some of those studies by screening a patient, initiating some sort of an acute insult, having some sort of a period of recovery when the pain stimulus reaches a particular intensity, moderate or severe usually. We will then dose the patient. We evaluate the response over a single dose and then we terminate assessments either after the dosing interval is over, which is generally 8, 12 or 24 hours, or at the time that the patient requests their first rescue analgesic.

[Slide]

There are compelling reasons why pharmaceutical sponsors have not gone down the path of efficacy evaluations in the multi-dose arena, and I would like to address these and propose some

potential solutions.

[Slide]

There is no doubt that there is no growing request for data. I recall that even at the Vioxx advisory committee meeting there was discussion of the availability or relative lack of multi-dose data in the dossier. There have been increasing requests from both Division 550 and 170 for such data.

I think the challenge here is, if I can just be frank and I guess this is for the record, that our collective rhetoric perhaps outpaces the actual science of drug development. In other words, our methodologic ability, to echo what Dr. Laska was saying, to actually tease out some of those differences is not always there.

In order to address this issue of multi-dose analgesic evaluation from an efficacy perspective, we need to ask ourselves precisely what our objectives are. Are they to establish efficacy? Are they to demonstrate effectiveness? Are we trying to establish dosing frequency? Are we trying to prospectively test a draft package insert? Or, are we merely trying to provide some sort of supportive safety data in a perioperative

setting where perhaps patients might be critically ill and otherwise compromised?

[Slide]

Here are some of the challenges to evaluating these drugs in acute pain. The first issue, and this has been alluded to earlier, is that the natural trajectory of acute pain is such that, whether treated or untreated, for the most part it diminishes. To be sure, and Dr. Katz referred earlier to thoracotomy patients or lumbar laminectomy patients who may have somewhat long-term pain. To be sure, some patients may have a longer trajectory, but a majority of these patients have a relatively short trajectory. So, this introduces an issue that most analgesiologists have called assay sensitivity.

We are also faced with a reduced duration of hospitalization. A significant number of patients after major surgery are home within four days to a week's time.

There is also a growing trend towards surgical techniques that reduce surgical pain. For instance, hip arthroplasty, as is currently being conducted, requires substantially less postoperative opioids than perhaps 10 or 15 years

ago and this presents a bit of a challenge.

Furthermore, patients will sometimes refuse to consent to multi-dose placebo controlled studies. It is one thing to convince patients to do a single-dose placebo controlled study, but to tell them you are going to repeatedly be give placebo over the next five or seven days presents a bit of a challenge.

We also have this issue of data contamination when you give rescue analgesia, and we have a problem in terms of availability of trained analgesic observers or nurse raters. This is a very specific discipline requiring an exceptionally well-trained individual who truly understands analgesic methodology, and there is a real shortage of such folks. Your most senior study coordinator usually wants to work the day shift so you have 72 hours more to go beyond that to evaluate the patient.

[Slide]

I would like to suggest some proposed approaches without getting too prescriptive. Some of these have really been spurred through discussions with Division 550 with Dr. Witter and Dr. Simon and others. One option clearly is to use

active controls, with the Division's prior consent. That is certainly one possibility to consider.

The other option is to use what I call pseudo placebos. So, these would not be placebos but would be perhaps ultra low dose of an approved agent, to allow us to get some assay sensitivity.

Yet another option, and this was discussed previously by Dr. Laska, is to use rescue analgesia as an endpoint. This has been used successfully but only with a modest degree of success in the past.

We can also integrate rescue and pain assessment data, and there are some techniques available for that. Of course, because of the shortage of trained study coordinators, we can perhaps consider doing serial assessments long term. We can use recall instruments to assess pain.

[Slide]

The rationale for integrating rescue and pain scores to come up with some composite scores is given on this slide, and I am going to be brief here. Traditional studies have tended to discard rescue after the first dose. The issue is that rescue tends to confound our analgesic evaluation.

Furthermore, rescue differentially confounds the analgesic response. David Silverman, for instance, has suggested a rather elegant but simple approach to integrating rescue and analgesia scores.

[Slide]

Alternative approaches that are available involve the use of recall instruments. We know that recall, at least among analgesiologists, is viewed as somewhat suspect but we, and others, have shown and have published data demonstrating that recall is actually quite sensitive. We have done studies where we have looked at recall in orthopedic pain and other models, and we think that this allows you perhaps to conserve on the resources that are a problem in multi-dose studies.

[Slide]

The last potential option that one ought to consider is rescue analgesia as an endpoint. I believe it is a potential endpoint. It does have some risks because the variability is not insignificant.

[Slide]

These are data that were presented in 1998 at the Arthritis Advisory Committee in the review of rofecoxib submission. As you can see in this

particular study, over day two to five there was a difference between placebo and rofecoxib in terms or rescue consumption. It was a one tablet per day difference. Now, whether this is clinically meaningful is a separate issue but it certainly provided some assay sensitivity in an attempt to look for differences.

In summary, the methodology for multi-dose efficacy evaluation is not quite cooked; it is not established. I think there are some possible options that are available, but we need to understand that there are some compelling reasons why single-dose evidences have formed the primary basis for efficacy evaluation. None of these techniques can meaningfully, in my opinion, answer questions related to the time course of effect and dose response. Those questions, and they are critical questions, need to be addressed in single-dose efficacy evaluations. Thank you.

Further Discussion of Criteria for Chronic Global Pain

DR. FIRESTEIN: Thank you very much. At this point Lee has asked us to revisit our discussion of the proposal for the criteria to obtain a chronic global pain indication. Just to

remind people, there are two essential issues. One is that for such an indication the proposal was that three separate models would need to be explored, and in each of them there would be three separate domains that would have to be all positive.

So, what we are going to do now is actually go around the table and get people's opinions on those issues. I would ask that people restrict their comments to two minutes or less. Please don't feel obligated to use the entire time because there are about twenty of us and it will take quite some time if we wax poetic.

I will go ahead and start and then people can take various and sundry pot shots at my comments, either amplify or deny them.

DR. ELASHOFF: I am still unclear on the question.

DR. FIRESTEIN: The question is what do the individual members feel about, number one, what the criteria should be for a chronic pain indication, with the initial proposal that there be three separate indications explored in order to get labeling for chronic pain.

DR. SIMON: Global chronic pain indication

with three areas of etiopathogenesis that would have to be studied with three domains as co-primaries in replicate trials.

DR. FIRESTEIN: So, those are the two separate issues that we should comment on. Does that clarify that?

DR. ANDERSON: But what are domains?

DR. SIMON: To remind you, they were patient global, function and a pain score. It is just in chronic pain. I know we have just talked about acute pain but we didn't get enough clarity yesterday for us to know exactly what you all felt about our proposal.

DR. FIRESTEIN: We were appropriately obtuse. So, I will start and then we will just go around the table. For introductions we went to my left and this time we will go to my right.

There were a number of other proposals that were also made with regard to the number of indications. First of all, I think that the bar should necessarily be high for a global chronic pain indication. The question whether it should be two, three, four or five indications is really not well defined by evidence-based medicine but, based on opinion, three doesn't sound like a lot and four

sounds okay and five sounds like a lot. So, by process of elimination, four sounded reasonable to me.

The other issue is whether or not you need replicate trials for a global pain indication. It seems to me that the indication is global pain, not the individual models. So, for instance, a confirmatory trial would not be a second OA trial but a second trial in another indication, preferably different mechanism, and I think there needs to be considerable care with regard to choosing how one selects the different models, making sure that there is adequate representation from multiple mechanisms--neuropathic pain, musculoskeletal pain, cancer pain, etc. So, from my perspective, it seems to me that a single trial with more indications makes sense.

With regard to the domains, the main issue is that function may not necessarily be a reasonable endpoint for some of these indications, as was pointed out yesterday, and I think there needs to be some flexibility in endpoint selection. Pain is obviously going to be the more important one and function may be less important in certain patients where strictly comfort is all that

matters.

So, why don't we move off to the right? Dr. Brandt? Was that clear enough?

DR. BRANDT: Fundamentally, I think I agree with Gary. The complexities in the science that drives chronic pain, as we heard yesterday, I think are very significant and it makes it hard to reduce this in terms of a limited number of models of disease states in which a drug shows efficacy to be comfortable that that truly gives enough information for a global pain indication. So, I am more comfortable considering pragmatics. I think it would be reasonable.

I think we regard to the outcome measures, certainly pain, certainly patient global, and I think that you have to look at function in terms of the specific disease state that is more relative to certain diseases than it is to others, as we heard. But I think the greater breadth that would be provided by demonstrated efficacy in four disease states for chronic pain has appeal to me, and perhaps more than looking at three times with the six-pack.

[Laughter]

DR. KATONA: Looking at the issue from the

pediatric point of view, for the chronic model it will be very difficult to recruit enough patients since out of the four proposed models really the only one which could be found in children in great numbers is the cancer pain. Children have no OA, very rarely low back pain, a low incidence of neuropathic pain. So, I think the study is going to be limited. The acute model I think is very important in children. So, those two will have to be concentrated on.

As far as efficacy, I think we always rely a lot on the adult trials and I think we definitely will do the same. However, I think the PK studies, the dosing schedule and especially the safety are going to be extremely, extremely important in children. So, I think those are going to have to be conducted and these have to be long term. Thank you.

DR. ABRAMSON: I would maybe take a slightly different position at least from Ken and Gary on this. I mean, chronic pain is a very broad term. Although it is clinically a very important issue, the name of the term itself is like the 1899 Merck Manual of Hepatology or lumbago and I think we have to be careful in setting a bar for a

broader indication that the elements within that indication are robust in the way that they are looked at from the term etiopathogenesis that Lee used.

Therefore, whether a global pain indication requires three, four or five individual etiopathogenic syndromes, I think the bar for each of those syndromes has to be as high as it would be for anything else that a drug is getting approved for, namely, two replicate pivotal studies for example.

When you talk about domains in these studies, the domains may vary within the syndrome you are looking at, whether it is neuropathic pain, low back pain, osteoarthritis pain, etc. So, clinical outcomes, meaningful clinical responses, things that you might tag on to look for mechanisms of pain will vary within each of those.

So, I would make the argument for keeping the bar very high for any individual entity of the individual syndromes that need to be looked at, recognizing that fibromyalgia is different from low back pain and the musculoskeletal indication for example.

Then, whether one gets for marketing

purposes a more global indication will depend on three, four or five very highly rigorous standard replicate studies that would have been required for independent registration.

DR. FIRESTEIN: Lee, would you just comment on whether or not this would change the bar for individual indications? In other words, that is a separate issue I think.

DR. SIMON: No, in fact, the bar, as we have described it in my earlier discussion, for any one indication with two replicate trials with three domains is obviously open to discussion based on which domains, but we would like patient global pain and a functional domain. It is particularly applicable to osteoarthritis but it may not be applicable to all of them. So, that would not change an individual indication issue.

What we are really discussing here is, is that high bar too high for the global chronic pain indication? And, we each have our opinion and that is what we are waiting to hear.

DR. WITTER: I just want to add a thought, and I think Dr. Katz brought it up yesterday. As you think about this, I mean, we are interested in labeling that makes sense to you as clinicians and

also to your patients. So, were we to construct chronic pain, the big claim, you know, I think you need to think through your current repertoire of medicines and ask if they should be able to reach that hurdle. If they do, then what implications does that have for whatever claim structure we might set up because would we be creating something and everybody would get it and may not have what we had hoped down the road. So, I think maybe you want to think about that as well.

DR. MANZI: I think when I was thinking about this the one assumption here that is probably true is that the number one biggest problem probably in the U.S. is that we under-treat chronic pain, more than abuse of medications or over-treatment. So, with that in mind, I said what would the advantage be of having a global indication more than industry incentive in some way? What advantage to the patient?

I guess from that perspective, I actually would presume that a global indication may open the door for a broader application of some of the potential medications in patients with chronic pain.

With that in mind, I would say what are

the downsides? The downsides may be that it is not as effective in certain disease states or that perhaps in certain subpopulations it may not be safe. I think those are clear concerns.

With that in mind, I guess my perspective is that I might actually consider lowering the bar a bit and say is it really safety issues and efficacy that we are worried about, or do we really want to open up to our patients the availability of a broad range of potentially helpful agents for treating chronic pain?

With that said, this is arbitrary but I would say I would go a little lower with perhaps the three entities not having to capture every pathophysiologic mechanism for pain because I am not sure that is even possible, obviously, keeping the individual rigor that the FDA does already with each of those entities. So, I think I would favor more a slightly lower overall bar to get a global label for the reasons that I mentioned.

As far as the domains, I agree with the previous speakers that I think you have to a priori determine which domains are relevant to the disease state that you are looking at and decide what the success is in each of those and not make a standard

requirement across the board for each population.

DR. KATZ: I feel more comfortable articulating some general principles relevant to this discussion, rather than just throwing out a number of five, three or something like that. So, I don't know if my comments will help you in any way but I will go ahead and take my two minutes or less anyway.

First of all, there has been a great debate as to whether giving an overall categorization for acute pain, chronic pain, or what-have-you, is appropriate. My feeling is that the opioids have taught us that it is possible to have a class of drugs that are broad spectrum analgesics for just about all kinds of pain. So, I think that the notion of a broad spectrum analgesic does have construct validity.

Number two, I think the opioids have also taught us that just because a drug has broad spectrum applicability in acute pain, chronic pain, it doesn't mean that it is going to work for all subcategories or all populations or all people. I think that is fine and it should not dissuade us from giving a broad sort of labeling, although it would be nice if we had some way, through the label

or otherwise, to educate physicians that just because a drug has a broad label doesn't mean it will work for everybody and it doesn't relieve them of their responsibility to manage their individual patient or different disorders.

I think acute pain as a category does have construct validity and I think chronic pain as a category does have construct validity too. It seems to me that in order for something to be called a medication for chronic pain, it needs to work for neuropathic pain as a broad construct and also for musculoskeletal pain because drugs that work for musculoskeletal pain may not work for neuropathic pain, and vice versa. So, it is inconceivable to me that something could be called a medication for chronic pain without working robustly in both of those different categories.

So, I wouldn't see it possible to label a drug for chronic pain unless one could also label it for neuropathic pain broadly and one could also label it for musculoskeletal pain broadly, with whatever robustness of evidence one would need in each of those individual subcategories.

We have just had a meeting for a whole day and talked about neuropathic pain and what sort of

trials would be necessary for that. People have thought that you would need a six-pack or more just for peripheral neuropathic pain, let alone chronic pains. That is a big discussion and I am not going to try to summarize it all here, but I think it is important to just say that you have to be confident of neuropathic pain before you get to the point of chronic pain.

In terms of the issue of replicate trials, personally I find it much more useful to see different trials in different disease entities than in the same entity. For example, two identical replicate trials in osteoarthritis don't help me nearly as much as one good trial in osteoarthritis and one good trial in some other kind of musculoskeletal pain like low back pain or rheumatoid arthritis, or something like that. I think that is where the information comes in. So, personally I would discourage replicate trials and, if you are looking for a broad categorization, then try to get as broad an experience as possible of disease entities within that category.

Lastly, in terms of the issue of the requirement for the three co-primaries, my experience suggests to me that that is an

absolutely wrong approach. I think it is obvious that if a drug reduces pain but does not necessarily improve function, quality of life or whatever, it is still an analgesic.

On the other hand, I think that those are very, very fundamentally important secondary outcome variables that will differ from disease to disease and can also help us understand the meaning of the primary and borderline cases or unusual cases. I think the data should definitely be collected. It should be required but not as co-primaries for developing analgesics.

DR. ANDERSON: I actually agree with quite a lot of what Dr. Katz said, although I disagree about the domains. First, I didn't like the idea of this global indication at all because I just don't think a single drug can do it all and also retain function. Also, it seems to me that it would be abused in the sense of, you know, you had all your three areas or even six areas where you showed it worked it would be used in many more where it might not work at all or might be unsafe.

So, I think that you should just stick with what you have at the moment, which is for any particular indication, pathogenesis area or

whatever, you have to have two trials, perhaps with a different disease.

I think that the three domains are all important. Okay, this is an analgesic but it is more than an analgesic. You know, for an analgesic which is just for acute pain, then, okay, pain is the only outcome that matters. But for an analgesic that is for chronic pain or long-lasting pain, then it is not much use unless the person can have function unless you are talking about terminal illness where there is no hope for that. But I think that we would want to use these drugs in cases where people want to retain and improve function. So, function, patient global and pain score I think are equally important and should all be kept and be required.

DR. ASHBURN: I am an anesthesiologist who has left the OR to take care of patients who have chronic disease over long periods of time. So, as a result, I am used to having conflict within myself.

[Laughter]

I think that this is one of the areas where I have mixed feelings. In a global area I think it is really important to recognize that

individuals who have complex chronic pain disorders require more than one medication. They frequently benefit from polypharmacy with medications targeted towards specific issues and specific individual patients. They frequently have depression; they frequently have sleep disorders; frequently have anxiety. They also have social issues that need to be addressed by cognitive behavioral therapy. They also have physical dysfunction and require activating physical therapy. To a certain degree, it is almost disingenuous to think that one medication could be useful as a global indication for chronic pain.

The other thing that even makes it more difficult in that area is that pain management physicians and physicians in general tend to be enamored with the use of unproven techniques in this patient population. I think that that poses some concern with regard to safety.

On the other hand, six well-controlled trials for the indication seems to be an extremely high bar. Drilling down to the specifics, I am a little bit worried about the specific definitions of group as far as how you define, how you group patients. One concern that was already brought up

is how would you study children, and for essentially orphan children who have chronic pain in these areas. Clearly, designing six well-controlled clinical trials that include adequate numbers in children would be extremely difficult. Do you do it by mechanism? Do you do it by cancer? We have already heard discussions that patients who have metastatic cancer don't necessarily have one etiology of their pain but frequently have multiple ones that are working simultaneously, and is that a meaningful patient population to study? Or, do you do it by body location, which also is fraught with all sorts of problems?

My concern is that if you set the bar too high companies will go for a narrow indication, which may be appropriate but, on the other hand, a narrow indication will lead towards less data on safety in different patient populations, which I think would be very helpful in guiding use.

With regard to a patient global indication, I think that this is something that probably ought to be required but I have a concern about it being used as a primary endpoint to determine approval. I think having six positives

is very, very difficult. Also, I don't know that the patient global assessment is well defined in the literature, and whether or not that assessment tool, which has become very common, has been validated in a meaningful and appropriate way and is used in a uniform and consistent manner.

Lastly, most of the function scales have multiple different measurement tools and they have to be well defined with regard to how you would affect function. The usefulness of a tool will vary by patient populations. So, it is possible that you will be offered different function assessment tools for different patient populations and you will not be able to combine that in a meaningful way. Again, with pediatrics there is very little data on validated disease-specific measures of health in children with pain, and even less data on children at the end of life. As a result, children are again going to be orphaned.

An alternative is to require the use of validated, as best one can, disease-specific measures of health specific for the population to be studied in each individual trial and use that data, not necessarily solely for determination of approvability, but use that to inform the label.

Thank you.

DR. ELASHOFF: I don't feel well enough informed to comment on the issue of how many separate indications one might make or what they would be. However, I do feel that each one going into that should have sufficient information. So, I feel very strongly that you should have replicate studies.

In terms of the outcome domains, probably each indication is going to need somewhat different ones, but the whole issue that I am concerned about is that all this needs to be extremely carefully defined before the study is started or, perhaps even before you talk about an indication for a specific area which things ought to be measured. the whole issue of exactly how one is going to deal with multiple co-primaries on a statistical basis, what you are going to do about alpha levels what the implications of this are for power, you will probably need to look very closely for each indication at how correlated these things are because that is going to have a great deal of influence on the powering of the study. If they are very highly correlated you are in essence only asking for one of them. If they have very low

correlation, then you may well need bigger sample sizes.

The other thing that wasn't put into the question, although some people have mentioned it, is that I think the safety requirements, the safety information that you would need if you are going to have a global indication should be far greater than for any single indication.

DR. FIRESTEIN: Dr. Farrar, you are up.

DR. FARRAR: I guess from my perspective, understanding that no drug is going to be perfect and that every drug is going to fail at something and that FDA approval is being used more and more to limit payment for therapies by insurance companies, I am in favor of a global indication to allow me to use medications in patients for which there is good clinical trial evidence that they work but which may not have been submitted to the FDA for formal approval, which is really very often driven by costs and marketing considerations.

As such, I think it is reasonable to think of a global indication. In fact, I would favor two trials in syndromes which are clearly neuropathic and would also request that those be in separate entities but clearly neuropathic, and two trials in

what are clearly somatic pain, also two separate entities as being the bar for efficacy.

In addition, since patients really are the defining factor in terms of whether a medication works or not, I think that the global outcome is exactly the right measure provided it is done correctly, and I think it can clearly be done incorrectly. By correctly, what I mean is that it is supported by several other outcomes that are all going in the same direction. To have a global outcome that is by itself I think would be incorrect.

In this setting, however, the most important issue and the thing for which the bar needs to be set the highest is safety. If the drug is going to be used or potentially used in a wide variety of patients, it needs to be shown to be safe in those populations, in specific, the elderly and children. It may be hard to find enough children to demonstrate efficacy in all of these areas, but if I know that it is going to be safe I would be willing to try it, and maybe clinical trials that are done outside of FDA approval will help to guide my therapy.

Lastly, I would like to suggest that

perhaps there needs to be a different study that is called perhaps a labeling study. We look at dose in a Phase II trial, but maybe we need to look at dose in Phase III(c) or perhaps even in Phase IV to help us answer some of these questions that have been raised in terms of whether a 50 percent response time is the appropriate dosing schedule if it, in fact, limits our use of the medication. In actual fact what we are talking about is limiting the use as opposed to providing real benefit in terms of the guidance for use. So, those would be my suggestions.

DR. BORENSTEIN: My thoughts on the subject have to do with trying to follow the clinical situation with the clinical setting. If we are going to have a chronic pain indication on a general basis, those situations for an individual neuropathic pain versus low back pain versus even osteoarthritis may not be quite the same. My hope would be that the FDA would allow studies to be done that could show potential efficacy that would mirror the clinical situation. Now, it may make it a little bit more difficult because the trials may have a different look to the patients that would be admitted and things of that sort. But it would

have greater applicability to what the clinical situation is.

So, whether that would be three or four settings where it would follow what would be happening in the clinical situation, that would make it much more applicable. So, this idea of either having multiple drugs and adding or withdrawing would then be allowed so that a trial for osteoarthritis might look different than one with neuropathic pain versus one with low back pain, but would still be accepted and how many would be needed, whether that would be two of each in neuropathic and somatic versus three, I think would still need to be decided.

I think also very important is the idea of safety and that the studies be done at least long enough for us to get a handle on how these agents would be used in these clinical situations. I think that is very important because it is all well and good to have a single drug and see whether it is safe but in the real world many patients are on three, four or five different drugs. They are hypertension drugs; diabetes drugs. And, it is the interaction of the new agent with the other ones which makes it, once again, clinically applicable.

So, I think the closer we can get to the real world and still do good science would certainly be quite useful.

The last point I would make regards the domains. I think a global assessment is clearly very important, but I think as an analgesic, we want to be sure that patients are achieving pain relief and that should be the primary outcome of studies. But every study should look at patient satisfaction and global outlook. So, I think those two at least. Then, in the appropriate setting how that is affecting their daily function and using the appropriate outcome measure to measure that would once again be important. But, once again, I think it is the clinical situation, as close as we can get to it, the greater will the impact will be of the information which is actually observed from these studies.

DR. STRAND: Well, I would like to perhaps give a little bit of a preview to what I was going to say this afternoon, after lunch. The group that I led at the NIH breakout meeting finally decided on five domains that they felt were essential as a minimum number of domains to be assessed in clinical trials of chronic pain. They were pain;

patient global; some type of measure of physical function or health-related quality of life, a generic measure of health-related quality of life and adverse events.

So, what we are really talking about here I think is that these need not necessarily be co-primaries. As has been done in other diseases, and I am not trying to shove this into the rheumatoid arthritis model, one could ask for any number of these five domains assessed by different instruments to show improvement without the others showing deterioration.

We could perhaps elevate patient global to something like a health utilities measure, which is more like the way the patient would weigh all risks and benefits from the intervention in terms of their pain and assess what they think of it.

Certainly, we talked about physical function and belabored the point that it doesn't work in metastatic cancer pain. I would simply argue that what we need to be doing is looking at the instrument. There are plenty of different instruments that would assess domain of some type of function--the ability to perform activities of daily living, the ability to even get out of bed,

whatever. They can be disease specific even down to the type of cancer that there is. So, I think there always is some instrument that would help in the clinical setting that we are looking at the pain.

Clearly, we have to ask about pain. A reason to look at a generic measure of health-related quality of life, besides economic assessments which might be important in noon-malignant types of pain, would also allow us to compare interventions across different kinds of pain. If we are talking about doing, say, three different models or four different models of chronic pain, somatic, musculoskeletal, or inflammatory as I would like to think of it, versus neuropathic.

Adverse events are obviously quite important and that was, of course, the fifth domain. In terms of the fact that these domains would not be closely related, if they are combined in some type of a responder analysis that should decrease the sample size quite significantly. It certainly is true with rheumatoid arthritis. In terms of saying that perhaps both the global and the pain measures, whatever they might be, have to

be required as improved and then the others must not show deterioration, or whatever, that is another way to make sure that the domains that everyone thinks are most important are specified. But it also makes it a lot easier than requiring that any three domains be co-primaries which is very difficult.

Finally, not to do any of this that isn't evidence based. I have been a part of predefining responder analyses on the basis of consensus with there being no data, and those are fraught with very much of a likelihood of failure, as Jane Elashoff has mentioned. But it could be done based on looking at data in Phase II with the product and then defining a responder analysis based on the data dredging from the Phase II studies.

DR. MCLESKEY: I would like to reiterate what I said basically yesterday, that I think we are all in this together. Our purpose, as I believe I mentioned yesterday, is to advance the practice of medicine and how might we best go about doing that

The concern that I expressed yesterday, I will reiterate today, and that is to study a new agent in three different models of disease, each

studied in a replicate fashion; each having three co-primary requirements that all have to hit in order to obtain a claim is, in fact, a high hurdle, perhaps too high a hurdle, perhaps a hurdle that you simple cannot get over. I am just concerned that if industry feels that it is such a high hurdle that it can't be achieved then that might, in fact, stifle innovation, which is the antithesis of what we are all about.

So, I just restate that again. I hope that I am reflecting adequately what industry in general feels, but it seems to me that the hurdle that has been proposed as a possibility seems a bit high and potentially challenging to a degree we can't meet.

Another issue, and it has been raised by previous panelists around the room, is that some of those co-primaries may actually be inappropriate in certain models of disease and, therefore, maybe those co-primaries need to be reexamined and reduced a little bit in their importance in certain circumstances. Also as was previously mentioned, the question of validation of some of the tools also potentially deserves a closer look.

The discussion yesterday regarding

multiple alternatives that has been reiterated today reminded me of a an advisory meeting that was held a couple of months ago, which Gary had mentioned earlier. It was a discussion of neuropathic pain and there were multiple possibilities mentioned at the time, one of which I will just reiterate for this group today, those who were not in attendance, because I haven't heard this particular possibility alluded to yet. As a suggestion, it was that one method or one model disease could be studied in replicate and then other models of disease studied not in replicate but in single form, sort of a combination or merge of the two different proposals. At that meeting, I heard mentioned that we might do a replicate analysis of one model and then look at maybe two other models of disease in a single study format to justify a broader claim.

Just as an aside, Lee, I would like to compliment you for mentioning yesterday and then highlighting again today the fact that you are proposing a subsequent meeting to examine these kinds of issues more closely, more carefully, perhaps in a more focused way in the presence of the academic community, the presence of the

regulatory community and perhaps a more meaningful presence from the industrial community as well, with representatives with a more substantial presence at that occasion. That is reassuring certainly to the industry members in the audience today.

As an aside also, I think some of the industry people would also like to be reassured, if that were possible, that the arrangements that are already under way and the commitments that have already been made will, in fact, be honored as these new guidance proposals are development and in process, some reassurance there would be appreciated, I know, by some in the room.

Also, just as an aside or perhaps as a commentary, some of the industry people have come up to me during the breaks and they are reflecting on the following, and that is the issue of idealism versus realism. There are many physicians and healthcare providers at this table in practice; there are many in the regulatory agency; there are many in the industrial organizations and sponsors that are in the room today and all of us know, as has been mentioned by many of the clinicians at the table, the variability in patients and the

variability in their circumstances. It is that variability that makes some of these trials so difficult to accomplish and complete in a fashion that would satisfy the proposal that is before us today.

That is why I am concerned that the hurdle might be set too high. We just must not lose perspective of the variability in patients and in their situations and in their circumstances which would make it very difficult to hit on all of the targets that have been proposed.

DR. FIRESTEIN: Before we move on, I would just like to remind people to please keep their comments to about two minutes, and let's try to answer the specific questions that have been raised. Dr. Max?

DR. MAX: Regarding the models, I agree with Dr. McLeskey that people are going to want to do replicate trials in one condition anyway to get the drug on the market. It would make sense to me. I would rather have a broader representation of diseases and I don't need any more replication. So, whether the number would be two and one, plus two additional conditions or three additional conditions, I would recommend that the FDA do a

careful economic analysis, and if you could get more conditions without killing the wonderful engine of industry, I would make it five trials, if not four trials, and you can figure that out.

I think in each condition you should try to either make it relatively homogeneous mechanistically for clinical criteria, or at least allow the information to be there. For instance, if you study cancer pain, mixed cancer pain means very little mechanistically. We should be able to look at bone pain separately and, similarly in back pain, the people with root injury are different from those with central back pain. So, try to use the clinical criteria to allow some mechanistic inferences.

Regarding the issue of the three proposed co-primaries, I again disagree with that. I think that pain should be the primary outcome. I agree that a global outcome and function are important things to measure but they should be secondary outcomes and, obviously, if over the pattern of studies globals deteriorate and function deteriorates there is something wrong with the drug and it won't be approved. But I would make pain the only primary. And, I think general chronic

pain is a great idea as it will drive the science forward.

DR. DIONNE: Well, I have very little experience with chronic pain so, presumably, I don't have the basis for an intelligent opinion but that hasn't stopped me before.

I just wanted to reiterate the concept of some sort of a data-driven regulatory practice for analgesic drug development in this particular question that might take the form of a meta-analysis of the existing drug classes that are generally accepted for chronic pain, be it tricyclics and NSAIDs, and look back and see if there is enough evidence to support the application of these criteria that are being considered prospectively when we look at the evidence that exists for drugs that have been studied for 50 to 100 years. Then, on the basis of that we might determine that the standard is too high, too low, if it doesn't actually apply to drugs that have already been approved, and then make the subjective evaluations that have to be made about the prospective criteria at least on the basis of the data for the drugs that are already out there.

DR. WOOLF: I must admit, I am concerned

about this notion of there being a global chronic pain analgesic in the absence of evidence that such a drug exists. I think that is the key issue. This needs to be evidence based. I am worried that we don't know which trials, whether they be three or five, in which conditions are going to be predictive of whether any drug is going to be effective across a wide range of different chronic pains.

So, the issue to me is how happy are we going to be living with an analgesic that has a global pain indication and, yet, is not effective in subcategories or different diseases? If we don't have a basis yet for predicting which of the suitable trials, whether it be low back pain or fibromyalgia or age-related neuropathy, it is pure guess work as to which of these we can select and how many to try to come to an assessment of whether any individual treatment is going to be effective across a wide range of conditions.

The other issue that hasn't been discussed yet is in these trials are we looking for placebo-controlled trials or active comparators? If so, since they are going to be so different what would the active comparator be if you are going to

compare fibromyalgia versus neuropathic pain in the conduct of these trials?

MS. MCBRAIR: I too am concerned about a global assessment. It seems early on and what I would really like to see us do is a really good job with each one of the indications or diseases or health problems and be able to give the very best guidance to the practitioners that are using these medications and to the patients. I think we need to focus on that first before we go towards a global assessment.

As far as the domains, I think they are all important based on the individual health problem. I do think patients need to be able to function if they are supposed to, and that is the goal of the medication in part. Certainly in rheumatoid arthritis, if we are just covering the pain we may not be addressing the inflammatory process and that needs to be paid attention to as we are looking at these individual situations. But I think the domains are very important to the people that we are trying to serve.

DR. WOOD: It is getting late. I agree with much of what has been said before, particularly by Dr. Abramson. I also agree with

what Dr. McLeskey said, that there are worries about having multiple primary endpoints and merging these into a composite endpoint rather than just having your primary endpoint being the reduction in pain which is, after all, the indication we are looking for.

On the other hand, a global indication seems to me to go beyond the science. If you think of other areas, we don't give global indications to improvement in cardiovascular health. We say cholesterol agents do one thing; beta blockers to something else; ACE inhibitors do something else. All of these drugs, in fact, produce mortality but we have a recognition about the specific indications for their use to reduce that mortality and that seems appropriate here; it is just that the science isn't as far advanced.

The one thing that has not been discussed that I would want to put on the table is that it seems to me there is an underlying assumption being made up till now that all our studies are going to come out positive in a global indication. What are we going to do with studies that come out negatively? Never mind how many positive studies you need, how many negative studies do you need?

Does one negative study immediately knock you out of the park? I mean is that it? That you can no longer get a global indication?

I would be particularly concerned that that is going to give rise to gaming of the system. You know, I think we can reliably expect that we will hear about all the positive studies. The negative studies may not be presented in this room. So, I think the idea that somehow all the studies will come out positive and really all we are arguing about, as Bernard Shaw said, is the number is unrealistic. Some are going to come out negative. And, I think there is a big danger for industry in going for a global indication because, clearly, if you go for a global indication and one of your studies comes out negative you are dead in terms of a global indication. There is a possibility that one of your competitors may come out with a study that is negative and that is then used to undercut your global indication.

So, I think there is a risk in that and I think we should be cautious about extending to indications for which we don't have obvious data to support them.

DR. CALLAHAN: Well, I think Dr. Woods

made a very good point about if there is a negative indication. So, based on that, I would like to say I would like to see two replications of whatever indications, and the numbers I think would depend on sort of the feasibility within the company in terms of how many indications they could look at. Clearly, you need to look at different types of mechanisms within that.

In terms of the domains, I think pain should be a primary outcome, not have the co-primary, but I would like to see some sort of disease specific function included, as well as patient global. Then, I very much like the idea of a general health-related quality of life so that they can be compared across conditions.

DR. CUSH: There is a benefit to going late; you get to listen to everybody else's ideas and be swayed by them. I will back off. I was very much in favor of this when it was first presented and I would say I am against it.

[Laughter]

DR. FIRESTEIN: I am going to have to go around the table again now, so be careful!

{Laughter]

DR. CUSH: I think that there is an issue

regarding under-treatment of pain, but I think that doesn't rest with the lack of available options or drugs that could be labeled as globally effective therapies. I think that rests more with poor education and poor understanding of pain and pain control. I think if you look at drugs that we might call sort of global drugs, widely used drugs, broad-spectrum antibiotics, while they may have been helpful there has also been a certain degree of misuse, and the problems that that may have arisen from that I don't think were anticipated.

When we look at our arthritis drugs, we have drugs like methotrexate and disease-modifying drugs. They tend to be used globally, sometimes outside of indications because we don't have options. Sometimes that is done because we understand the mechanism of disease. Sometimes it is done quite blindly and quite stupidly, and with no apparent effect and maybe with great expense or maybe toxicity. I think that there are drugs that are out there that are being used in this manner currently, drugs such as the COX-2's and narcotics, are basic globally used pain medicines. Currently they are used in a way that basically forces the physician to be intelligent and understand the

mechanisms of disease and what is going on with the patient, and also act as an advocate on behalf of the patient to go for those indications and write letters to explain why this is indicated.

So, you know, would a global indication actually help a payer, an approver of drugs that they may not be indicated for? So, would they actually approve the use of a new, novel pain medicine for phantom limb pain, acute gout or visceral pain associated with losing to the Yankees? I don't know.

[Laughter]

I still think it forces me to have to still write those letters to get these drugs approved, and for this reason I would say that we should not have this indication.

I will close by just saying I think we have an issue of nomenclature here that was raised yesterday by Dr. Ashburn. The whole use of words "acute" and "chronic" are a little bit disconcerting and I think we should try to maybe redefine the terms we use and maybe go for things such as short-term therapy or long-term therapy. In this instance, general global pain indication is a bit too obtuse clinically and unrestrictive to be

useful. Thank you.

DR. SHERRER: I am last but I didn't change my mind. So, some of us can stay steady. While it is true that we do, in fact, use medications that are on the market with restrictive indications broadly, nevertheless, as a clinician, I think it would be very useful to me in prescribing to know that a drug has utility across different types of pain. If the studies were useful and really are showing me that, for instance, if you do osteoarthritis and low back as two of your models I am not so sure that you are looking at different pain. On the other hand, if you look at cancer bone pain and you look at diabetic neuropathy and you look at OA, you probably are looking at different pains and it would be very useful for me to know that that has been demonstrated.

In terms of looking at the domains, I am one of those who believes that we need to look at the total impact of the drug as an outcome. So, I would favor looking at least at three, if not four of them. I think pain is useful but the total impact of a drug is even more useful to my patients. In fact, that is why some won't take

certain pain medications, because of the side effects, because of their effect on quality of life. So, I would use several of those, and most important to me would be pain, would be patient global and some appropriate assessment for the particular disease of function or quality of life.

One thing I haven't heard that I would like to bring up, and maybe it would be a secondary, is steroid sparing because I think that in certain chronic pain disorders where steroids are an important part--I said steroid sparing, opioid sparing--many patients are very concerned about opioids and so are we, and if a drug spares opioids, that would be very important to me.

DR. FIRESTEIN: We are done. We have gone all the way around the table. So, we will break for lunch and we will reconvene at 12:55, which means we will start at 1:00.

[Whereupon, at 12:05 p.m., the proceedings were recessed for lunch, to reconvene at 1:00 p.m.]

A F T E R N O O N P R O C E E D I N G S

DR. FIRESTEIN: I am happy to introduce Dr. Vibeke Strand, who is going to talk about responder index, a model.

Responder Index, a Model

DR. STRAND: Thank you, Gary. We have been more or less talking around this topic for the last day and a half, and perhaps we should have started sooner with this discussion.

[Slide]

What I would like to do is basically present to you a discussion that was started at the last NIH-FDA meeting on pain. Just to point out something that we have talked about before, responder analyses have face and content validity. They do allow the assessment of multiple domains. They probably could better help us categorize analgesics.

They should also help facilitate comparison of efficacy across products and disease populations and indications. I think in analgesia, as in rheumatology, most of our patient populations are quite heterogeneous and this would help considerably.

This might or might not lead to a tiered

approach in label indications as has been done in rheumatoid arthritis but really has not yet been done otherwise. The precedent, as we have talked about previously, is the ACR responder criteria in rheumatoid arthritis.

[Slide]

Jim Witter pointed this out to you this morning. it is a model for other responder analyses. One could say that the two criteria here, which are tender and swollen joint count, could be required in a responder analysis in pain, for instance, whatever assessment of pain could be required and perhaps also the patient global assessment could be required. The others could be included.

One of the things we do know is that it is probably too stringent to require all components of a responder analysis to be improved. It is possible to choose the majority of them to be improved. It is also possible to indicate that the remaining ones should not be deteriorated.

If we want to talk about a definition of no deterioration, however, we have to allow that statistical definition to account for test/retest variability, which we have alluded to before in our

discussions around changes in visual analog scales.

[Slide]

The strength of the rheumatoid arthritis guidance document is that it has had a proven track record and since its inception we now have six products approved for the treatment of rheumatoid arthritis, some of them just for the signs and symptoms, as in the COX-2 products, but many of them now for improvement in signs and symptoms in either 6 or 12 months and then inhibition of radiographic progression at 12 months, and subsequently improvement in physical function without deterioration in health-related quality of life over 2 to 5 years. In this case it has been over 24 months.

These outcomes have all been achieved in single protocols using prespecified outcome criteria, whereby the first outcome criterion must be satisfied statistically significantly, p less than 0.05. Then one may look at the subsequent, in sequence, criteria, provided each one remains statistically significant without taking a p value correction. That is a very valuable way to look at multiple different aspects of a disease and how it affects the disease population.

[Slide]

When we had this breakout session at the workshop, in May, the definition for the workshop--and I am not saying that is a definition we have been working on today, but the definition for chronic pain was randomized, controlled trials of at least three months duration in pain of at least three months duration, regardless of the underlying cause. That was simply taken as a definition so we could have the discussion we were going to have.

We agreed in that discussion that we would not specify specifically different diseases. We agreed that maybe there might be some differences specifically for chronic cancer pain, but for the purposes of the discussion we would not distinguish.

[Slide]

We were considering musculoskeletal indications such as rheumatoid arthritis, osteoarthritis and low back pain, as we have talked about in the last two days, also fibromyalgia, neuropathic pain, the examples being diabetic neuropathy, post herpetic neuralgia, trigeminal neuropathy. For cancer pain, we agreed that it

wouldn't necessarily be for a three-month duration in terms of trial and that we would be thinking about rapidly progressive disease and adjust intervention as the disease progresses which is, of course, a very important thing around cancer pain.

[Slide]

We agreed to select the domains regardless of the clinical indication; that we would consider the available instruments and whether or not they were validated and whether or not they had been validated in pain trials; just that they had been used in previous randomized, controlled trials but not necessarily in pain; and whether they were disease specific or generic was sufficient.

The point really was that the outcome measures in rheumatology clinical trials, the OMERACT international consensus process has actually helped to define the ACR responder criteria, and is helping to define responder criteria in osteoarthritis, but the first decision is around the domains to be used, not the specific instruments, and that there is some flexibility around which instruments might be utilized to satisfy each of the domains.

[Slide]

We did believe, however, that the strength of choices in terms of domains was based on multiple available instruments and our own prior clinical experience. So, the choices, as they were thrown out and written up, were pain and we talked a lot about the multiple different measures of pain that probably should be important to be included in a given trial under a single domain, including the patient global assessment; including the assessment of rescue medications; and time to treatment failure--all of these which we talked about this morning.

Suffering was suggested as a domain, as was pain relief, a disease specific measure of improvement and/or physical function and/or health-related quality of life was proposed. So was health-related quality of life, and we have been throwing around the term quality of life. I think it is important that we specifically mention that it should be health-related quality of life in all the way health affects you. Because, certainly, political circumstances, economic circumstances and the presence or absence of food and money are not part of health-related quality of life but certainly are part of quality of life.

Patient global assessment, adverse events and specifically how they are perceived by the patient which is something we are not very good at in clinical trials; we usually trust the physician to report those adverse events and often not with very much input from the patient, other than that the complaint has been offered. Damage, whether it is due to the disease or its treatment, and specifically indicating that it is irreversible, and economics.

[Slide]

After a relatively brief series of discussions, we came up with the final vote the first time when everyone was allowed to vote on basically three parameters: Unanimous decision for pain; an almost unanimous decision for a disease specific or a disease relevant measure. We have been talking a lot about physical function but, as I said to you before, I think it can be basically perceived as a disease relevant or specific measure of either function or health-related question. Health-related quality of life as a generic measure was an almost unanimous decision. Patient global and adverse events followed.

So, this was felt to recommend a minimum

core set of required domains, and that other ones could certainly be added but if we were to speak about trying to do a responder analysis, these should be the components to be considered at a minimum.

[Slide]

We have talked a lot about defining improvement in pain, but I think the point we are all trying to get at is defining improvement multidimensionally. We know that patients experience pain and they report pain, but they report it specifically as they feel on the day they are reporting it. So, if they are forward filling their diaries, it is based on how they are feeling that day. If they are back filling, it is also based on how they are feeling that day.

One of the important things too is that their expectations of what they can do and what they should be able to day change according to how their pain is. So, if they have already had significant pain relief their expectations have changed and become even greater than they were when they, for instance, first entered the study and were suffering considerable pain.

What we are trying to do, obviously, is

separate the experience of pain from functional impairment and disability which may or may not occur because of the pain or follow the pain. We want to separate physical impairment from disability. It is important, I think, to use individual responder analyses because it allows us to define responder, non-responder. We don't have to impute data. All cause dropouts before the endpoint are then considered non-responders. Therefore, from a statistical analysis it can be a more robust analysis. I think it is important that we use both disease specific or disease relevant measures as well as generic measures.

[Slide]

Something to quickly point out is that disability is really in the eyes of the beholder. It is, of course, age and gender appropriate. It is important and pertinent to the work, the family and the social setting. But, in fact, someone who has had cerebral palsy since birth and is wheelchair-bound may not perceive themselves as being disabled even though we would certainly consider them to be far more than just physically impaired.

The other part of it is that impairment

may be due to pain or it may be due to structural alterations, and functional limitations are certainly something that we can measure. There are arguments about disease specific or disease relevant measures of physical function and how accurate they are in that those of us who are rheumatologists often note that our fibromyalgia patients are far more severely impaired than our rheumatoid arthritis patients. But, by and large, if we can choose the right types of instruments we can usually find some type of a valid report that is consistent with the other self-reports that the patient may offer.

[Slide]

One of the other things about a global assessment is that it is probably much more important to ask the patient in all the ways that your pain is affecting you, including its treatment--how are you doing today? When we talk about visual analog scales for patient global assessments, we always talk about how are you doing today, this moment? The other part of it here is to make it a global assessment and to include sort of the risk as well as the benefit as an important thing in terms of the patient assessment of the

pain treatment.

Now, a transition question can probably be equally sensitive, in other words, how are you compared to when you first started taking this medication? That may well get to the same point.

The other point that is quite useful is that health utilities which are used for economic measures are single reports sometimes, questions or several questions around how patients are doing in terms of what their perception of perfect health would be. A health utilities index or the EQ5D can be given. It is a simple questionnaire that the patients can fill out. Or, one can ask the patient to report, by a feeling thermometer, how they are doing in terms of perfect health and death. That looks very much like a visual analog scale vertically.

[Slide]

We have talked a lot about minimum clinically important differences. We consider them to represent changes which are perceptible to patients and are considered clinically important and meaningful. When they were first started in the OMERACT process we used patient query as well as a delphi technique. Then they were demonstrated

to be consistent with patient global assessments of improvement or patient global assessments of how they were doing.

In fact, when we determined the proportion of patients with clinically meaningful improvement or clinically important improvement, this gives us a much more interpretable result than, in fact, trying to say, okay, this many patients had 50 percent improvement in pain or this many patients had 30 percent improvement in pain.

[Slide]

If we think about this, we have now noticed that changes in disease specific or relevant measures of function and health-related quality of life that have been statistically related to much or very much improvement in patient global assessments, either by visual analog scale or Likert have given us very consistent values across OA, RA and fibromyalgia, and I will show you that briefly.

[Slide]

Briefly, measures of chronic pain include a lot of different things. There is the brief pain inventory, the McGill pain questionnaire, all of these others. Perhaps one of the more important

new ones is the treatment outcomes and pain survey which was developed as an add-on to the SF-36 and has been shown to be very useful in cancer pain, as well as some other non-malignant settings of pain, chronic pain with multidimensional therapy.

[Slide]

The faces rating scale we have talked about before. We talked about using a visual analog scale that is not anchored. This one actually combines a Likert scale of more or less 7 with a visual analog scale of 10 and is sort of the example of what not to do at the same time to get sensitivity and specificity, which is why I chose to show this slide because I, myself, would be very confused about which face to combine with which number.

[Slide]

Talking about MCID, one of the nice papers published by Dr. Farrar, sitting at the table, is looking at the pain intensity numerical rating scale and comparing that to very much improved in patient global assessment.

These are 10 placebo, randomized control trials of Pregabalin, which is not yet approved, but this has been published in Pain 2000 for

diabetic neuropathy, low back pain, fibromyalgia and OA. So trials across different indications of chronic pain have shown that the relationship of much and very much improved in PGIC and pain intensity by numerical rating scale is very consistent with reduction of 30 percent or two points in the pain intensity scale.

This is really interesting given the wide variety of disease states here, and this is regardless of the baseline pain scores in these patients. So, a robust MCID definition.

[Slide]

If we look at other measures of physical function and health-related quality of life in chronic pain, I just want to remind you again that the top survey here is meant to look at changes in health-related quality of life in individuals over time, which is different from the generic measure of health-related quality of life, the SF-36, which I will come back to in a minute, and one other measure that is an HRQOL measure in pain is the MPI which specifically looks at psychosocial role functioning but omits work-related activity. Finally, cancer-related health-related quality of life has been looked at a lot on the BPI, the brief

pain inventory, but that has not been validated in non-malignant pain.

[Slide]

Generic health-related quality of life measures go back as far as the sickness impact profile which is, in fact, considered not to be a very popular instrument because it implies to the patient that they are sick.

The Nottingham health profile is also an older measure of HRQOL and not particularly popular. A very popular one is the SF-36 which is expanded over the SF-12. It is designed to measure health-related quality of life in large groups and across different disease states. It has problems if it is being used as a single measure of HRQOL in pain states or in arthritis states because there is a limited assessment of upper extremity function, as well as upper extremity pain and facial pain, and does not differentiate well between low back pain and upper body pain.

The WHOQOL is a new instrument, but with 100 questions it has fallen out of favor. There are some shorter version. The EQ5D is widely used in Europe.

[Slide]

Disease specific measures of physical function and/or health-related quality of life include all of these. We have called them disease specific. People like Jim Freis, who developed the health assessment questionnaire, prefers not to call it disease specific because he believes it can be used across many disease states as well as aging, which is not a state of disease, as he wants to remind me. So, I have chosen to also call these disease relevant measures.

Clearly, the WOMAC is something that is a very good one for osteoarthritis of a knee or a hip. There are others, as well as some for the hand which are being developed. We talked about Roland-Morris and Oswestry. There are some for geriatrics and, of course, a variety of ones for cancer.

[Slide]

What I would like to do very quickly is just show you some examples of how these measures interrelate in rheumatoid arthritis, osteoarthritis and fibromyalgia.

[Slide]

So, if we look at rheumatoid arthritis, we talk about the health assessment questionnaire

which has now become widely used in randomized controlled trials in rheumatoid arthritis. It is a measure of physical function with 20 questions. It also accounts for when patients use aids or devices to perform these activities.

[Slide]

The SF-36, as I mentioned to you, is validated and widely used. It has been validated across multiple cultures, many disease states. There exist gender and age specific norms for multiple populations, both in the U.S., Canada and northern Europe and other countries. Then, it has eight domains as well as a physical component score and a mental component score. It has been shown in RCTs to show change in as short a time as four weeks, probably sooner than that.

[Slide]

The physical domains are physical function role, physical body pain, general health. They are combined positively into the physical component score which then negatively also weights the mental domains of vitality, social function, emotional and mental health. So, positive changes here are weighted positively here against the positive changes in these domains, which are negatively

weighted for the mental component score. The mental and physical component scores are based on normative data only to a total of 50. Therefore, they can show less change. And, if you are looking at a disease like rheumatoid arthritis where the predominant change is in the physical component domains, then one is not going to be seeing much improvement in mental domains because they are weighed against by the improvements in these.

[Slide]

What we have learned from the various trials is MCID for the HAQ disability index is a score 0.22 improvement. For the SF-36 it is about 5 to 10 points in domains. For the physical and mental component scores, 2.5 to 5 points.

[Slide]

So, if I look very quickly across some clinical trials in rheumatoid arthritis you can see, with the leflunomide Phase III trials across all three studies, with methotrexate and sulfasalazine the mean improvement over two years exceeds MCID almost to twice in all treatment groups.

[Slide]

If we look at the ATTRACT study, and this

is HAQ disability index over two years, again we see that in the placebo group it does not quite reach MCID and is about twice that in all of the active treatment groups.

[Slide]

Similar types of improvements in the ERA trials with Etanercept versus methotrexate.

[Slide]

If we go back to look at the U.S. study with leflunomide and methotrexate, which was the first to show that the SF-36 was sensitive to change in rheumatoid arthritis, you can see that based against age and gender matched U.S. norms the patient population had significant decrements in all domains of healthcare quality of life, but particularly physical function, role physical, bodily pain and vitality. As we know, patients perceive their health-related quality of life differently, and one can see the changes here in the active groups actually are within MCID for almost every domain, with some deterioration in placebo.

[Slide]

If one then goes forward, we see that these are the baselines for the treatment groups

and these are the age and gender matched norms, then treatment with leflunomide and methotrexate, in fact, just about bring health-related quality of life up to a normative population level. That is probably a very meaningful change and it certainly does equal MCID in many of these eight domains.

[Slide]

There is similar improvement infliximab in the ATTRACT trial. These are the two of the physical domains. If we look at the PCS and the MCS we see that there is very significant decrement in the physical component score at baseline, almost two standard deviations from the U.S. norm, and treatment over one and two years brings it to within one standard deviation of the U.S. norm. As we might expect, the MCS was not that different from expected, and it could not show a great deal of improvement based on the large amount of improvement in the physical domains. Nonetheless, improvement is shown.

[Slide]

This is the median improvement in PCS score with the ATTRACT trial showing the same type of a picture, with placebo showing not much improvement.

[Slide]

This is the early RA trial, again showing baseline for the PCS, about two standard deviations below the U.S. norm, and improvement to approximately one standard deviation from the U.S. norm with treatment.

[Slide]

So, I think you can see from this that basically improvements in HAQ disability index, in other words the disease relevant measure of physical function and the generic measure of health-related quality of life appear to be very clinically meaningful, and that there are consistent values for MCID across these instruments. We are showing that improvement in a disease relevant measure is highly correlated with a generic instrument, and the generic instrument is useful because we can compare it across different disease states for an economic basis, but also to try and understand improvement, for instance as we might when we are looking at chronic pain indications.

[Slide]

Quickly, lets look at osteoarthritis. The WOMAC is the disease specific measure in OA of the

knee and hip. It reflects physical activities that are most affected by the osteoarthritis. It is composed of pain, five questions on joint stiffness; two questions on physical function which dominates the instrument of 17 questions out of a total of 24, and is scored either by a zero to 4 Likert or a zero to 10 VAS scale for each question.

[Slide]

So, what we have found out looking at the COX-2 trials with both celecoxib and roficoxib is that basically, using a Likert scale for the composite total WOMAC score, MCID was about 10 points and was different according to the domains because they had more or less questions. If one uses the VAS scale for all of the questions, then we see very consistent MCID for each of the domains of about approximately 10.

[Slide]

This is what this looks like in the composite scores of WOMAC in clinical trials of celecoxib versus placebo and the active comparator, naproxen. Here is MCID.

[Slide]

If we look at it for rofecoxib using the primary outcome question in the physical function

subscale we see again that improvement is evident and exceeds MCID considerably.

[Slide]

If we look at the improvement in the SF-36 with rofecoxib and we compare it to age differences in the U.S. population, we can see that there is considerable improvement in the mental domains as well as the physical domains, but the largest improvement is in role physical.

[Slide]

Similarly, if we look at the changes with celecoxib in the SF-36 in the trials that I showed you previously, you can again see that MCID is reached in many of the domains, particularly the physical ones.

[Slide]

This actually translates again towards improvement that approaches the U.S. norm. This is the U.S. normative population and these are the final scores with the different doses of celecoxib and naproxen and placebo.

[Slide]

So, again, we see clinically meaningful improvements. We see that the MCIDs are consistent across agents and patient populations in this

disease, and that improvement in the WOMAC correlates with the generic HRQOL SF-36 measure.

[Slide]

I don't have outcomes for fibromyalgia, but I do have interesting consistent relationship at baseline between pain, sleep disturbance and fatigue. These are all patient reported and they are highly correlated either by a pain diary or a sleep quality diary or multidimensional assessment of fatigue, a well-known fatigue instrument. And, this is whether it is done by a numerical rating scale that is ostensibly recorded daily in the diary or a visual analog scale that is done at the office visit weekly. It has been shown that the high baseline scores indicate impaired sleep. Significant fatigue, we know that our fibromyalgia patients think of themselves as being very physically impaired, and these correlate with low scores in SF-36, particularly role physical, bodily pain and vitality domains; poor sleep quality by the MOSA sleep, high fatigue and also more anxiety than really depression.

[Slide]

In terms of cancer, there are a lot of different instruments that would be useful in

trials of cancer pain, and they can be the FACT-G or FACT that is a P for prostate or any one of the cancers that you want to look at. The same for LASAs which can also be done for symptoms of chemotherapy as well as for symptoms for cancer or pain. The same kind of thing for the FLIC. Basically, there are all these different instruments that can be used and, again as I mentioned to you before, the TOPS has been developed and validated in cancer pain, among others.

[Slide]

Since the TOPS was defined as an extension of the SF-36 it has been a very useful instrument and it really does show change in individual patients over time.

[Slide]

So, the appropriate domains, based on what we discussed at that particular breakout session and as a recommendation to this group, would be that pain would be included as a domain. There are many instruments. We have talked about looking at different ways of assessing pain. Perhaps we can get away from some of our old visual analog scales and Face scales.

A disease specific or disease relevant measure of health-related quality of life and the ways that the disease affects you in your day to day activities could be used, or one could use the TOPS which is much more generic. When it is relevant to whatever the disease is, other measures could be looked at. They do not necessarily have to be included in the responder analysis.

I think you can see that the health-related quality of life measure SF-36 as a generic measure has turned out to be very useful and sensitive to change across a large number of types of diseases; and some way of asking the patient how they are doing in terms of risk/benefit in terms of the treatment as well as the pain; and finally adverse events, which we haven't talked about, might be subsumed under this global assessment if it does include the treatment as well as the pain.

[Slide]

Certainly for acute pain we probably don't need a measure of health-related quality of life, as we have discussed, and certainly we can talk about all of these. We do want to remember time to treatment failure and rescue medications as being

part of something that needs to be assessed in the pain domain.

[Slide]

When we go to subacute pain or pain of two to five days, or whatever the definition is that is less than chronic pain but more than one day of pain, it would appear that these different domains would be equally relevant. We can show changes in SF-36 over a very short period of time. Again, it might be useful to use the TOPS or to use a disease relevant measure.

[Slide]

In fact, again Dr. Farrar has published a very nice paper on cancer-related breakthrough pain, acute pain. This was in a study of oral transmucosal fentanyl citrate, which ultimately was not approved. But these were 130 patients who were naive to the study drug, many episodes of pain, and the differences in pain scores between the episodes which did and did not yield adequate pain relief. Again we see MCIDs for pain intensity difference and maximum total pain relief of about 33 percent. Again, the same kinds of changes in terms of absolute pain relief and sum of pain intensity differences of 205 points in a Likert scale, which