

jam

AT

DEPARTMENT OF HEALTH AND HUMAN SERVICES  
FOOD AND DRUG ADMINISTRATION  
CENTER FOR DRUG EVALUATION AND RESEARCH

ARTHRITIS ADVISORY COMMITTEE

**Guidance for Industry**  
**Clinical Development Programs for Drugs, Devices**  
**and Biological Products**  
**for the Treatment of Rheumatoid Arthritis (RA)**

Volume II

Wednesday, February 5, 1997

8:07 a.m.

Gaithersburg Hilton  
620 Perry Parkway

MILLER REPORTING COMPANY, INC.  
507 C Street, N.E.  
Washington, D.C. 20002  
(202) 546-6666

jam

Grand Ballroom  
Gaithersburg, Maryland

PARTICIPANTS

Michelle A. Petri, M.D., M.P.H., Chairperson  
Kathleen Reedy, Executive Secretary

MEMBERS

Steven B. Abramson, M.D.  
David T. Felson, M.D., M.P.H.  
Felix Fernandez-Madrid, M.D.  
Matthew H. Liang, M.D., M.P.H.  
Daniel J. Lovell, M.D., M.P.H.  
Harvinder S. Luthra, M.D.  
Leona M. Malone  
Frank Pucino, Jr., Pharm. D.  
Lee S. Simon, M.D.  
Barbara C. Tilley, Ph.D.

CONSULTANTS

Karyl S. Barron, M.D.  
M. Clinton Miller III, Ph.D.  
Patience H. White, M.D.

FDA

Wiley A. Chambers, M.D.  
Frederick Miller, Ph.D.  
Kent R. Johnson, M.D.  
William Schwieterman, M.D.  
Jeffrey Siegel, M.D.  
Lisa Rider, M.D.  
Janet Woodcock, M.D.

C O N T E N T S

PAGE

Introductory Comments:

Michelle Petri, M.D., M.P.H.....4

Conflict of Interest Statement:

Kathleen Reedy.....5

Open Public Comment:

Mark Waltrous, Ph.D.....9

Ken Seamon, Ph.D.....14

David Smith, Ph.D.....17

Introduction to Document and Discussion of RA Claims  
Structure:

Janet Woodcock.....18

Signs and Symptoms:

Janet Woodcock.....62

Pro/Con Debate re: Dropouts:

Kent Johnson and Jeff Siegel.....89

Function/Quality of Life:

Kent Johnson.....116

Structure (X-ray/Other):

Jeff Siegel.....141

Major Clinical Response:

Kent Johnson.....154

Toward a Data-Driven Definition:

David Felson.....156

Complete Clinical Response and Remission:

William Schwieterman.....170

jam

Overview of JRA and Subsets:

Patience White.....194

C O N T E N T S (cont'd)

PAGE

Application of Pediatric Rule, JRA Claims  
Structure, JRA Claims and JRA Drug Development:  
    Lisa Rider.....201

Equivalency Trials:  
    Wiley Chambers.....246

Safety Analysis and Phase IV:  
    Frederick Miller and William Schwieterman.....265

Conclusions and Summary:  
    Janet Woodcock and Michelle Petri.....309

- - -

P R O C E E D I N G S

**Introductory Remarks**

DR. PETRI: Good morning. My name is Michelle Petri. We are here for the Arthritis Advisory Committee.

I would like to ask Kathleen Reedy to give us the meeting statement.

**Conflict of Interest Statement**

MS. REEDY: The Conflict of Interest Statement for the Arthritis Advisory Committee on February 5, 1997. The following announcement addresses the issue of conflict of interest with regard to this meeting and is made a part of the record to preclude even the appearance of such at this meeting.

Based on the submitted agenda for the meeting and all financial interests reported by the committee participants, it has been determined that all interests in firms regulated by the Center for Drug Evaluation and Research which have been reported by the participants present no potential for an appearance of a conflict of interest at this meeting with the following exception. In accordance with 18 United States Code 208(b)(3), a full waiver has been granted to Ms. Leona Malone.

A copy of this waiver statement may be obtained from the agency's Freedom of Information Office, Room 12A-30

jam

of the Parklawn Building.

In addition, we would like to note that Dr. harvinder Luthra's employer, the Mayo Clinic, has an interest in American Home Products. Lederle, a subsidiary of American Home Products, is the manufacturer of a competing product to Neoral, which is unrelated to the firm's competing product. Although this interest does not constitute a financial interest in the particular matter within the meaning of 18 United States Code 208, it could create the appearance of the conflict. However, it has been determined, notwithstanding this interest, that it is in the agency's best interest to have Dr. Luthra participate in the committee discussion concerning Neoral.

In the event that the discussions involve any other products or firms not already on the agenda for which an FDA participant has a financial interest, the participants are aware of the need to exclude themselves from such involvement, and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose products they may wish to comment upon.

DR. PETRI: Thank you.

jam

I would like to have our panel and invited guests introduce themselves, and we will start with Dr. White.

DR. WHITE: At the end of the table here is "W." I am always at the end of the classroom here. Dr. Patience White, chair of the Adult and Pediatric Rheumatology Divisions, here at George Washington University.

DR. TILLEY: I am Barbara Tilley, director of Biostatistics and Research Epidemiology at the Henry Ford Health Sciences Center in Detroit, Michigan.

DR. SIMON: I am Lee Simon, a rheumatologist at Harvard Medical School.

DR. SCHWIETERMAN: I am Bill Schwieterman from the Center for Biologics, Division of Clinical Trial Design and Analysis.

DR. RIDER: Lisa Rider, Division of Monoclonal Antibodies, CBER.

DR. [CLINTON] MILLER: Clint Miller, biometrician from Medical University of South Carolina.

DR. [FREDERICK] MILLER: Fred Miller from the Division of Monoclonal Antibodies, Center for Biologics, Evaluation and Research.

DR. WOODCOCK: I am Janet Woodcock. I'm a rheumatologist. I am the head of the Center for Drugs.

MS. MALONE: Leona Malone, consumer

jam

representative.

DR. LUTHRA: I am Harvey Luthra, chair of Rheumatology at the Mayo Clinic.

DR. LOVELL: Dan Lovell, pediatric rheumatologist, University of Cincinnati.

DR. LIANG: Matthew Liang, a rheumatologist from Boston.

DR. FERNANDEZ-MADRID: Felix Fernandez-Madrid, a rheumatologist, Wayne State University.

DR. FELSON: I am David Felson. I am a rheumatologist and epidemiologist from Boston University.

DR. CHAMBERS: I am Wiley Chambers. I am the acting director for the Division of Antiinflammatory, Analgesic and Ophthalmic Drug Products in the Center for Drug Evaluation and Research.

DR. BARRON: I am Karyl Barron. I am a pediatric rheumatologist and the deputy scientific director for the National Institute of Allergy and Infectious Diseases.

DR. ABRAMSON: Steve Abramson, rheumatologist, NYU in the Hospital for Joint Diseases.

DR. PETRI: Thank you.

We will start with the open public hearing.

### **Open Public Hearing**

DR. PETRI: There are several speakers who have

jam

already registered with us. I would like to encourage both as part of the open public hearing and the rest of this session the active participation of those of you in the audience.

The first registered participant in the open public hearing is Dr. Mark Watrous from SmithKlineBeecham Pharmaceuticals.

DR. WATROUS: Good morning. Thank you for the opportunity to comment on the guidance document.

I would specifically like to discuss Section 1B which is in reference to functionability and quality of life claims.

In the absence of cure, functional and health status measures have an important role in the clinical development --

DR. PETRI: Excuse me for interrupting.

DR. WATROUS: I'm sorry.

DR. PETRI: Could you just guide us to what page? I think that will help us keep up.

DR. WATROUS: I believe it is page 4, Section 1B.

In the absence of cure, functional and health status measures have an important role in the clinical development programs of rheumatoid arthritis. The committee's inclusion of such guidance statements involving

jam

these measures is commended and demonstrates a clear recognition of the importance and value of patients' perceptions of his or her health and the effects of treatment on those perceptions.

Specifically, I would like to thank the committee for the opportunity to comment on the following issues: the differentiation between functional status and health status/quality of life measures; the specification of specific instruments for functional status and quality of life claims; and the length of trials to support such claims.

Specifically referencing functional status in health status measures, outcome measures can be classified in the hierarchy of three domains, functional status, health status, and quality of life, with quality of life as the broadest of the three categories.

However, in many instances, these three domains are used inappropriately and interchanged. We fear that in the case of this guidance document, that may be the case. Such distinctions should be noted since functional status specifically references only physical functioning, whereas health status is a much broad concept representing physical, psychological, and social well-being.

Therefore, we would recommend that the document

jam

recognize a potential for two supportive claims, one being that of improvements and maintenance of functionability, the second being improvement and/or maintenance of quality of life.

In reference to the specific instruments referenced in the document, the field is growing and is in a phase of methodological and theoretical development as it relates to rheumatology. This is reflected in a few, but growing number of studies that are using these types of instruments to assess patients' perceptions of their disease.

As a result, there is currently no well-recognized "gold standard" measure of functional and quality of life for rheumatoids. Therefore, we would recommend that the agency be a bit more generic in their stance in terms of the selection of specific instruments.

Along these lines, we feel that the committee may be in a better position to make recommendations in terms of validation steps in the interpretation of instruments as they are developed, similar to the preceding section which outlines the necessary steps to support the signs and symptoms claim.

In reference to the specific instruments that are recognized in the document, the existing literature supports

jam

a reference for the HAQ where the HAQ has been proven sensitive to clinically significant changes in RA in a number of studies over a period as early as three to six months, and we would concur in this in terms of a support of a functional ability claim.

In reference to the SF-36, while this is a very well-validated instrument across a number of diseases, it is relatively new to the field of rheumatology. Original data by Ware looking at the instrument showed that it was most sensitive to changes along the physical function domains.

Most recently, at the ACR meeting in October, Dr. John Ware presented work of the sensitivity of the SF-36 within the clinical trial setting over a period of two to six weeks. Again, these data demonstrated that the instrument was sensitive to changes along, again, the functional domains and less sensitive to the mental health domains, questioning the additional or incremental value of this instrument over a measure such as the HAQ.

Also, along the lines of recommendations for specific instruments, we question whether this may not preclude the development and validation of additional instruments that we are aware of being developed currently.

Specifically, terms of timing of administration, as I alluded to earlier, it has been demonstrated with

jam

functional status measures, such as the HAQ and the AMES, that these are sensitive to change over periods as early as three to six months, and we would concur that such timings be done in conjunction with measures of signs and symptoms.

In reference to the timing involving an instrument such as the SF-36, validation work to date that we are aware of is the Ware study which shows a six-week period. One could hypothesize that changes in mental health domains would occur over a much longer period of time, such as one year to two years. However, there is no data that we are aware of to date that would allow one to make a recommendation along those lines. So we would encourage the committee in terms of recommendations on these instruments to also consider how data from these instruments should be interpreted to support such claims.

Finally, the last piece of Section 1B refers to the timing of signs and symptoms assessment in conjunction with functional and quality of life, stating that it could be within the same trial or having been previously demonstrated in other trials.

I would encourage the committee to consider that it is necessary to have both within the same trial; therefore, you having signs and symptoms data, as well as your measures of functional status and quality of life to

jam

support such claims. This will allow for interpretation of the clinical significance of changes within these instruments and also support validation of existing instruments as they are used in new trials.

Thank you.

DR. PETRI: Are there any immediate questions from the panel?

[No response.]

DR. PETRI: Thank you.

The next registered speaker in the open public hearing is Dr. Ken Seamon.

DR. SEAMON: Good morning. My name is Ken Seamon from Immunex Corporation. First of all, we would like to thank the FDA for their efforts to obtain outside comment and input in the development of this draft guidance. In particular, we would like to commend the rheumatology working group comprised of members from the Center for Biologics, Center for Drugs, and Center for Devices for working together to create this document.

We believe the guidance document provides appropriate requirements for assessing the safety and effectiveness of products for rheumatoid arthritis. However, we find the comments on the safety risks of biological products on the whole to be somewhat negative and

jam

not representative of current experience.

Biologics, and more specifically protein-based therapeutics, have been used to treat a variety of diseases and have an impressive safety record over the past 10 years. Various industry sources report that over 10 million people have received recombinant DNA-derived insulin on a daily basis. Over 1 million have received recombinant tissue plasminogen activator. Over 500,000 have received erythropoietin, and over 100,000 have received growth factors.

These products and other biologics demonstrate no unique safety risks that can be generalized to the entire class of protein-based therapeutics.

Protein-based therapeutics for treatment of rheumatoid arthritis function via a spectrum of different mechanisms of action. For example, they can exert their pharmacologic response by blocking cytokine receptors on the surfaces of immune cells, by binding and neutralizing soluble cytokines before they are bound to receptors, or by suppressing specific immune cells' ability to elicit a response. These protein-based therapeutic molecules have the potential to ameliorate the signs and symptoms of rheumatoid arthritis within a short period of time with minimal adverse events.

jam

The clinical data presented at the recent American College of Rheumatology meeting suggested many of the different protein-based therapeutic molecules may also be able to halt or slow joint destruction in patients with active rheumatoid arthritis more safely and perhaps more effectively than current therapies.

However, the section on special considerations for biological products makes a very strong statement about the potential safety concerns of these products. For example, on page 35 of the document, it is stated that, "The toxicity response curve may be highly unpredictable and potentially very dangerous, and include the risk of disease worsening."

We encourage the agency and the advisory committee to carefully evaluate the specific guidance with respect to biological products. Because each biologic exerts its effect through different mechanisms, each needs to be evaluated on its own merits. This should be based on a thorough assessment of its mechanism of action, its preclinical data, and its clinical safety profile, with the safety risks weighed against the clinical benefits of the product.

It is critical that appropriate safety data be acquired for all products that will be chronically administered to these patients. The agency has demonstrated

jam

a willingness to work proactively to make therapies available more rapidly for the treatment of life-threatening diseases. We hope that the agency and the committee will continue to work in this manner in the evaluation of other therapeutic products for the treatment of severely debilitating diseases such as rheumatoid arthritis.

Thank you.

DR. PETRI: Thank you.

Are there any immediate questions from the panel?

We will move on to the next registered speaker, Dr. Alan Solinger from IDEC Pharmaceuticals.

[No response.]

DR. PETRI: It appears he is not here. So we will move on to the next registered speaker who is Dr. David Smith from Hoffman La Roche.

DR. SMITH: Good morning. I would like to briefly raise a point for a possible consideration by the committee. I would like to refer specifically to Part 1 of the draft guidance document, Claims for Treatment of RA. This is on page 2, third paragraph.

This goes on to state that, "In addition to the traditional claim of improving signs and symptoms, other clinically relevant outcomes can be considered as label claims. We are specifically interested in the claim for

jam

prevention of structural damage. The guideline goes on to state that it is anticipated that under most circumstances, any of the additional claims will be approved only if there is adequate evidence to support the signs and symptoms claims.

We would like to have the committee consider the possible dissociation of these, and that given the strong medical need to prevent structural damage, ask the question if a drug could be approved for prevention of structural damage in the absence of an effect on signs and symptoms.

We would like to inform the committee that there are drugs currently in development with such a profile, and we believe that the final guidance document should make provision for approval of such a drug.

Thank you.

DR. PETRI: Thank you.

Are there immediate questions?

[No response.]

DR. PETRI: We all think your point is well taken.

Thank you.

I am going to now turn this part of the meeting over to Dr. Woodcock.

**Introduction to Document and Discussion  
of RA Claims Structure**

jam

DR. WOODCOCK: Thank you.

[Slide.]

DR. WOODCOCK: The purpose of this meeting today, the reason the FDA is bringing this draft guidance before the committee, is part of a process of developing a new set of guidances that will apply to drugs, biologicals, and to some extent to medical devices.

There was an existing guidance for drugs, but there were a number of compelling reasons that led us to work in a tri-center manner to try and develop a new draft guidance, and these are the following.

As we have already hear this morning, there are a number of rheumatoid arthritis treatments in development, and some of these are novel compared to traditional treatments. They include drugs, biologicals, and devices.

There was felt to be a need for consistency, particularly in the kind of claims that would be granted to products, no matter whether they were considered drugs or devices or whatever, and therefore, we developed this as a three-center document.

The prior drugs guideline did need to be updated, given many of the changes in outcome measures and so forth, and I will get into that a little bit more in a minute.

There were new outcome measures that had been

jam

proposed. I think there was some recognition that the current set of outcome measures that had been used had some shortcomings, and groups, including the International League Against Rheumatism, the American College of Rheumatology, and the OMERACT Group, were trying to develop various new outcome measures that could be used.

We recognize within the agency that the description of DMARDs, disease modifying anti-rheumatic drugs was not wholly satisfactory to describe the new agents that were coming forward and perhaps even to describe our current agents.

We all, I think, share a hope for better treatments for rheumatoid arthritis, treatments that truly can impact on the natural history of the disease, and we wanted to develop a claims structure that could recognize that and encourage it.

If I could have the next one, Rose.

[Slide.]

DR. WOODCOCK: Now, this guideline has already gone through -- this draft has already gone through a process. It was released as a draft on March 5, 1996, and we had a public workshop where we invited many people to come and comment on sections of this draft, and as a result of all the input we received there, the draft was

jam

substantially modified.

In addition, in July, there was a public workshop out on the JRA aspects of this draft, and we received a very lively session. We received a lot of comment on this, and we have modified this part of the draft in addition, in response to the comments.

Because of the many comments and many points of view around JRA that were articulated here, we are having a segment here today and we are going to try and bring some of the issues before the committee.

We have received a lot of written comments from academia and from industry, but people want more specifics in this document, and that is another reason we would like to discuss it today.

As everyone here knows, our FDA guidelines are not binding. They are not requirements. They are really guidelines. They are goal posts, and they represent what might be acceptable. However, we need some specific numbers in this document that people can react to or aim for, and we hope to consult the committee today on that.

What we hope to do is release this document as a final after incorporating the advisory committee comments. However, depending on the amount of comment we received today and the amount of recommended changes, we may, in

jam

fact, release it as a draft for comment one additional time, perhaps have the committee comment on it before it is released in final.

[Slide.]

DR. WOODCOCK: Now, the structure of this meeting, the FDA staff will present sections of the document in order to orient those, especially those of you in the audience who may not be aware of the structure of the document.

We will have specific questions that we will present about each section, but we are requesting comment from the committee, from the advisory committee, and also from interested members of the public here to the extent there is time on all aspects of this document. We really would like to have as much comment as possible.

[Slide.]

DR. WOODCOCK: Now, the structure of the document itself is the following. It starts out with a claim structure, a new claim structure for rheumatoid arthritis. That is in the very beginning of the document.

Then there are sections on RA product development. These are recommendations as far as what we feel would be good points to take into consideration in the preclinical development of these products, the early clinical development, and then we discuss possible trial designs for

jam

efficacy and considerations for safety evaluation of these products, especially taking into account the knowledge that the FDA has about the various products that have been developed and trials that have been done.

[Slide.]

DR. WOODCOCK: Following these recommendations, there are sections on special considerations, first for biological products, and we have already heard a little bit about that in the public comment, a section on medical devices, and then a substantial section on juvenile rheumatoid arthritis that really reflects the problems that we have, that there are very few agents approved or licensed for JRA specifically.

[Slide.]

DR. WOODCOCK: Now, I want to talk about the proposed claims structure first. This is a departure from what we have had in the past, and we think this is a very central part of this document.

I am sorry. I don't know it off the top of my head. I think it starts on page 2 of the document.

The first claim that is recommended is a reduction in signs and symptoms. Now, we will be discussing each of these claims in more detail immediately following my presentation. So what we would like to discuss at this

jam

point is here is the overall claims structure, are there things missing, are there things in here that shouldn't be in there and so forth.

The next claim we have already heard about is the functional ability and health-related quality of life claim. This is an explicit acknowledgement that RA is a symptomatic disease that limits quality of life and limits functional ability, and we want to be able to have a claim that reflects amelioration of that.

The third one is a more traditional claim. It is prevention of structural damage or joint destruction, radiographic joint destruction.

This is followed by a set of three claims, complete clinical response on medication which equals the attainment of remission off medication. These claims are very similar, except that one, in one that the patient still requires medication, and the other, the patient has remitted and is off medication.

The last category is also very similar, major clinical response. This claim, though, we need some help. We need help from any parties who can help us on this.

It was pointed out at our workshop in the spring that many patients with RA cannot achieve the criteria for remission because they have too much fixed joint destruction

jam

or joint problems that do not allow them to reach the ACR criteria.

We are trying to devise a claim that would acknowledge a substantial clinical response, a drug that is giving a very substantial response that could be measured in these patients, a response that is over and above what would get your signs and symptoms claim, almost akin to remission or complete clinical response in those patients with fixed deformities.

Could I have the next one?

[Slide.]

DR. WOODCOCK: So that's the claims structure I'm going to be asking you to comment on.

Now, the comparison to the extant or previous CDER guideline is the following. The CDER guideline had a signs and symptoms claim. This claim, as you will hear later, that we have in this guidance document substantially expands the ways in which you can attain the claim; in other words, the outcome measures, but the signs and symptom claim remains.

The structural damage claim was also in the previous guideline, although here, it's quite expanded as far as how you might attain it and so forth.

All these other claims are new claims, the

jam

functional ability, health-related quality of life, and the three related claims that relate to remission or substantial clinical response.

In addition, within this new guideline, any concept of disease-modifying anti-rheumatic drug or short-acting and long-acting, these concepts are gone from the document, and I think we would like the advisory committee's comments on those at some point.

We had found in granting claims to approved agents that drugs don't really fit in very well to these categories, and I think the efforts of a number of bodies over the past few years in trying to develop new categories to describe these illustrate the difficulties of having disease-modifying anti-rheumatic drug and so on.

So what we propose to do is to simply have this claims structure. Then the label can describe the time frames where the drug was observed to have its beneficial effect and so on.

May I have the next one?

[Slide.]

DR. WOODCOCK: Now, I hate to bring this up because it is very confusing, but I think the advisory committee needs to recognize and be aware that for all these claims, there are a number of dimensions that each claim

jam

could have, depending on how it is evaluated and what claim the sponsor is seeking from any one of those claims that we just went over.

For each claim, a product can simply claim to be effective. A product is effective in preventing structural damage. That would be the claim. A product is effective in treating the signs and symptoms and rheumatoid arthritis, but there are two other kinds of claims that are comparative claims.

One is the product can have effectiveness equivalent to some other named agent, to methotrexate. That is a somewhat different kind of claim than claiming your product is simply effective.

Or, you can claim that your product has effectiveness superior to a nonsteroidal, to gold, to whatever. That is another kind of claim.

[Slide.]

DR. WOODCOCK: Now, for the simple effectiveness claims, there are a number of trials that would support a claim of effectiveness, unqualified, noncomparative.

You can do trials where you show superiority to placebo and demonstrate that the product is effective in that manner. You can do trials that show the product is as effective as an active control. This is called an

jam

equivalency trial, but the sponsor may or may not be seeking an actual claim of equivalence, if you follow me.

You may be able to show that the product is effective by demonstrating it is more effective than an active control; in other words, it is superior.

[Slide.]

DR. WOODCOCK: The second comparative claim could be to claim equivalent effectiveness; that your product -- to put on the label that this product will give as much benefit, say, as methotrexate, put that right in front of the label, and then the sponsors could advertise, "This product delivers as much benefit as methotrexate."

To develop an equivalence claim is a little bit more substantial, obviously, than trying to just develop a claim that your product is effective. This requires head-to-head trials with the comparator.

A third arm, at least in part of the trial of a placebo, is very desirable, and we will get into the reasons for this, this afternoon, but sometimes it is difficult to tell whether both products are equally effective or they are equally ineffective.

As you all know, we sometimes see trials of RA treatments where they don't work any better than placebo, and if you did a trial like that and you had no placebo arm,

jam

you could conclude that the test product was as effective as the comparator product when, in fact, in the trial neither of them were effective.

This is something we have to struggle with. Obviously, in RA, we can't have long placebo-controlled trials. So it is something we still have to deal with.

In analyzing these trials, we use a confidence limit approach. We don't just look at the comparison of the means and say, oh, they were the same. We look at the confidence limits on each of the estimates.

The next one, please.

[Slide.]

DR. WOODCOCK: The third claim that sponsors may well seek in this environment, and we hope to see this -- we hope to see better treatments -- is we find treatments that are superior to existing treatments.

Again, this type of claim, seeking this type of claim requires that the sponsor pursue head-to-head trials with the comparator. In this case, you don't need a placebo arm because you are showing a difference.

We are recommending in this document that two trials be done to achieve this claim, both of them showing superiority, and we use regular statistical evidence of superiority as the analytical approach to these kind of

jam

trials, just as you would for a placebo-controlled trial.

[Slide.]

DR. WOODCOCK: Now, we seek some advise from the committee at this point on the claims structure, and the major dimensions of our request is, is the claims structure clear, can you understand what is being meant by each of the claims, is it accurate, does it really describe events in RA that are reasonable to have treatments for, is it complete.

We are particularly concerned that we received a number of comments about disability, progressive disability. I think we recognize that this, as a long-term sequela of RA, is really one of the most devastating consequences.

The claim isn't on there because we couldn't figure out how to structure the trials and the outcome measures, not because we don't think that that would be a valuable claim.

We suspect that if we could get it on, it would not be a claim that would be the first claim a product would be approved for. Products would be approved for more proximate claims, but perhaps could later be studied for prevention of disability because those would probably be longer-term trials. However, we do not have it on there because we could not imagine how to devise such a trial and an outcome measure.

jam

So those are the questions, along with any other comments, the committee would have on the claims structure that we would like some advice on.

DR. PETRI: Rose, if you wouldn't mind turning off the overhead.

Let's open the discussion among the panel. I would like to start by a question that I have, but which also came up in several of the industry communications to the FDA.

Obviously, many of these claims are actually subheaded under signs and symptoms and that you can't get the additional claims unless you have already met signs and symptoms. So those ones, I think, that would be subheaded would probably be functional, the complaint clinical response, remission and major clinical response.

I am not sure that they shouldn't just be subheads of signs and symptoms. In other words, there are different degrees of the signs and symptoms claim.

Let me ask the panel how they feel about having all of these as separate claims as opposed to subhead claims under signs and symptoms.

Let me ask Dr. Felson to start.

DR. FELSON: I have many, many things I wanted to comment on because a lot of the questions that Janet brought

jam

up are questions that the ACR committee is involved in, this process now for a number of years. We have thought about almost in exactly the same terms that you are thinking about.

We didn't think about drug approval. So I haven't really thought about signs and symptoms as sort of an overarching claim under which there might be subsidiary claims.

Let me comment on the disability issue because there are a number of facets that will speak to whether that is a separate aim or not. Let me comment first about that, and then let me comment about all of these different definitions of response that Janet put up as possible separate claims because I think those are two different issues that need to be addressed separately.

The core set and ACR improvement criteria include within them a measure of physical disability or physical function. I think we owe a debt of gratitude to the guy from industry who stood up and tried to help us distinguish between different concepts here. I think we have to be clear about the different concepts.

The ACR and ILAR and WHO core sets make quite clear that the core set measure here is one of the measurement of physical disability, self-reported, that is

jam

included in the HAQ. That is a measure of physical disability or physical dysfunction, and it is a component scale of the AIMS, of the SF-36, and there are a number of instruments that have been validated for use in rheumatoid arthritis and demonstrated to be sensitive to change that measure physical disability either as their sole point of measurement or as a component point of measurement, and our analyses of AIMS and others analyses of other trials suggest that physical disability is what tends to change most in rheumatoid arthritis in trials.

Now, should there be a separate claim for physical disability? In our analysis of trials and in others, physical disability improvement correlates greatly with two other measures in the core set and that are measured usually in rheumatoid arthritis, included in the core set.

One is the assessment of pain, which is closely correlated with physical disability, and the other is patient global assessment which is in the core set and in the original FDA measure. So it would be my strong suspicion that any trial which included the core set and demonstrated the improvement in patients in the core set would almost necessarily demonstrate improvement in disability because the two are closely correlated. It is not like a different claim.

jam

Anything that gets core set improvement or ACR improvement is almost always going to get disability improvement based on what is in the core set and how closely those things are correlated with one another and the fact that physical disability is already in there.

So I think, in heuristic terms, theoretical terms, maybe it is nice to have a separate claim that the company can come in with that, in fact, if you make it on what is called here signs and symptoms, and it actually includes disability, you are going to make it on disability, almost always. It would be peculiar if you didn't. It would be some analytic quirk. It would be because the definition, the requirement for disability improvement requires an area under the curve mean analysis, and the ACR improvement requires, perhaps, a point-in-time percentage improvement, and there was some reason for a difference between those, and that would drive a difference between the HAW result or the AIMS result or the SF-36 result and the ACR core set or ACR improvement result.

So I think they are essentially the same, and I think we ought to recognize that. There may be reasons, theoretical reasons, policy reasons, FDA reasons, and public reasons to talk about disability as a separate entity, but if we get improvement in signs and symptoms, we can get

jam

improvement in disability.

I wanted to comment, also, on major improvement in all of these other measures, but we can sort of hold there, if you want.

DR. WOODCOCK: Well, I think this is useful. Perhaps we have a definitional problem.

There is functional ability, diminution of functional ability on a day-to-day basis, but there is also long-term progressive disability, and I have heard a lot of commentators in the field say that treatment of signs and symptoms actually disassociated, and what we are talking about here is long-term disability claim, prevention of actual major functional status losses.

In multiple sclerosis, I think the Center for Biologics has approved something that prevents that long-term loss, permanent loss of physical function that we also see in rheumatoid arthritis, and that is what we don't think we are able to capture.

DR. PETRI: If I could just clarify that. I think we have to be very careful when we are talking about the structural damage and the acute inflammatory part of rheumatoid arthritis, and both things together, I think, go into the long-term disability. Obviously, the structural damage is probably not going to be reversible.

jam

David, did you want to comment more on that?

DR. FELSON: Janet, I agree with what you said.

You are sort of quoting the Ted Pincus and Fred Wolfe part of rheumatology that would get up and give a very compelling talk about the fact that short term we can see changes in tender-joint count sed rate, all of the other parameters that we are used to following in clinical trials, and yet, long term, over five years, over six years, over seven years, you can see those improvements transiently in those other things, stabilization of those other things, and then the patient's HAQ score continues to march down the road.

I am not sure that we can make clinical trials into long-term observational studies, and I am not sure we should try to. I think you correctly pose the question, but let me change the way you posed it a little bit.

Are we interested in a long-term outcome as a claim here, or are we interested in a short-term disability outcome?

If we are interested in a short-term disability outcome, you already have it in the ACR improvement criteria. If you are interested in a long-term, does this prevent that steady march in decline of functional capabilities of physical function? That is really a different question, and it requires a five-year trial or

jam

some really expensive long-term thing.

DR. WOODCOCK: That is not heard of -- unheard of, though, in other diseases.

DR. PETRI: Dr. Simon.

DR. SIMON: I would agree with David, but I would point out at this juncture that there is a little danger here of disassociating biology from actually observational science, whatever that is.

I am a little concerned that we perhaps have not included within this document, which I presume will last several years in the future given the effort that has gone into this, that yes, indeed, we may have drugs in the future that could re-create structure, that could put back in what we have lost; that, in fact, what we are really interested in is curing the disease, not just making people palliating pain and inflammation, and I think that is one of the key issues here that we do have short-term clinical trial experience in palliation of pain and inflammation, but we are now bordering or on the brink of a new era of being able to actually perhaps for the first time change the biology of the process.

In so doing, we have to figure out criteria to measure that. Well, the reality is that this long-term slow slope down, this slippery slope into disability is clearly

jam

related to structural abnormalities, and that we can halt structural abnormalities as one issue, which is a measurable phenomenon objectively, and if we can reverse them, they would be different and you would evaluate them differently under those circumstances.

I would argue that we need to be very clear as to how we will measure them. We heard an illusion to that as we thought about the disassociation of structural halting of disease, perhaps with a metalloproteinase inhibitor, for example, in some disease of cartilage, and just decreasing signs and symptoms where you could make somebody's disease better, but yet, they would still have pain and inflammation.

So it is hard to imagine that, but nonetheless, this document could potentially do that, and I think that is the issue that David has brought up.

I am a little nervous that if we are incredibly careful about trying to find a lot of observational things that we are going to lose the reality of what it means to change the disease biology, and we have to be able to address both in this document. I think your points are well taken about that, David.

DR. PETRI: I think Dr. Simon's point was also brought up in the open public hearing that the structure

jam

claim should be a stand-alone claim because it is quite possible that a drug would be developed that biologically could have a structural effect and have no effect at all on the acute inflammatory signs and symptoms.

I think my point, though, is that a lot of the other claims seem to be subsidiaries of signs and symptoms. Among the documents we got from industry was a lot of confusion about how many trials did you have to do for these different claims.

In fact, I think their very strong suggestion was that these could be subsidiary claims from a signs and symptoms effectiveness or equivalency trial.

DR. SIMON: That is really only because we are still using the paradigm of observational assessment. In thinking of these drugs only in the way we have had them, drugs or biologics, the reality is that we have to think about it in a new way, and we have to, thus, think about them perhaps not a subsidiary phenomenon, but they are perhaps stand-alone as well in a different world than the world we have been living in.

I would agree with you as to what we have had to date. I would agree entirely they should be subsidiary. I am just not so sure in the future, and I think we are going to hamstring us in the future in future evaluations.

jam

DR. PETRI: Dr. Miller?

DR. [CLINTON] MILLER: You argue about whether or not these claims should be stand-alones or not. It seems to me that we see a myriad of risk factors and a myriad of benefactors. This is clearly going to be a multidimensional decision space.

I don't think you can afford to make a decision on a single criteria that you are going to have to, as a group and as a professional group, make your judgments in that multidimensional space.

The structure, this claims structure that you are devising, insists on that. You can't do just one of those things at a time, and it is not impossible to do a multidimensional kind of decision.

We have the technology to do that. We have the designs to do that, and we also have the technology to do a better job of analyzing the kinds of experiments that are developed to show superiority or equivalence, et cetera.

I have a number of slides that speak specifically to that, and maybe this afternoon, when we get to that aspect of the implications of how to handle the data, when we do arrive at that structure, maybe it would be helpful for us to review those.

DR. PETRI: Thank you.

jam

Dr. Felson?

DR. FELSON: Michelle, would this be an appropriate time to talk about the other claims here, complete clinical response remission and major clinical response claims?

DR. PETRI: I think this should be open for a discussion of this entire session of the guidance document.

DR. FELSON: Let me suggest what I think these suggested claims represent. It is a tiered approach to improvement.

Any patient in a trial, given these different categories, could be characterized as having no improvement, ACR-level improvement, which is the reduction in signs and symptoms, actually. Let's call it that.

Then, the next level up would be something called major clinical response. Then, the next level up after that would be called complete clinical response or remission, depending on whether the patient is on or off medications.

So a patient entering into a trial could achieve any one of those levels. One of the questions is -- and Barbara Tilley and I had this discussion at lunch yesterday -- how many levels do you want.

It turns out, as Dr. Miller just implied, one can get more physical efficiency perhaps from using more levels,

jam

and that may not be unreasonable, as long as those levels, I think, are chosen with some attention to what is clinically reasonable and will have meaning to the clinician out in practice for which a particular approval is granted.

I do have particular trouble distinguishing between No. 4 and No. 5. I think it is artificial to distinguish between somebody on medication and off medication. We could probably debate for a couple of days what medication meant in rheumatoid arthritis, knowing full well that we all have patients who are on background medication.

I also could see companies attempting to come up with a remission version of this by defining medication on or off in a specific way that we might not necessarily feel comfortable with.

I think the idea of whether they are on or off medication is quite irrelevant. I think they should be defined as meeting remission based on ACR criteria.

Now, historically, that has been nearly impossible for anybody to get to in a clinical trial. I should tell you, though, I think we have all been to the ACR meetings recently. I think this is changing. I think there is secular improvement in treatment.

I think what is happening is that there actually

jam

are some patients I have seen in trials that are placebo-controlled and the patients have actually reached remission. There has been a percentage that is large enough to actually analyze.

So I think we need to be attentive to this threshold because I think we may actually see it. I wouldn't have said that five years ago.

Personally, I think the Nos. 4 and 5 ought to just be combined. I think that distinction is artificial. I think what we are interested in is how much improvement or how little disease they have remaining.

Now, one other thing, I think, we ought to talk about from a theoretical sort of design point of view here is that there are two delta measures here. One is reduction in signs and symptoms. A patient can improve by 20 percent. Another is major clinical response. A patient can improve by 50 to 60 percent.

Another in this list is completely different. It is remission. That is not the same. So patients entering a trial who may have very mild disease may have a much greater likelihood of going into remission, and patients with severe 10-year disease in some of the trials we saw yesterday, no matter what we give them, they will never go into remission.

In the context of a placebo-controlled trial, that

jam

may be unimportant because you can distinguish between the placebo and the treatment group, but a drug company or a biologics company going for a claim of remission will necessarily choose that group of patients which has a chance. Many of these patients don't have a chance.

That is one of the reasons why we didn't use remission when we developed the ACR criteria because we thought it discriminated too much against trials where salvage patients often entered.

DR. PETRI: David, aren't you actually asking that we know what sort of risk modifiers the patients have in clinical trials? We need to know rheumatoid factor positivity, shared epitope.

DR. FELSON: So what factors are there that would negate their chance of responding or of going into remission?

DR. PETRI: Yes.

DR. FELSON: Yes. I think that is right.

I think one of the major factors we know, and I will mention this later when we talk about major clinical response, is duration of disease.

People with early disease -- and we have done analyses of this and others have also -- people with early disease have a much better chance of responding to treatment

jam

than people with disease of 10 years or longer.

DR. PETRI: There is a comment from the audience.

Dr. Strand?

DR. STRAND: This discussion about whether claims are iterative and whether they should be combined or not leads me to suggest about looking at it a slightly different way.

I think we have to think about it from a time duration instead. We are looking at short term versus long term.

From the point of view of trying to develop therapies, we need to have, say, short-term outcomes which aren't a reasonable amount of time to show a difference or we are not going to have products developed.

For instance, once a signs and symptoms claim might be gained, could there not be continued work to then support a long-term improvement of a sort of longevity claim? So that, it would be, more or less, either Phase 4 work, Phase 3(b) work or some type of a continuation of what was already done.

I wonder if this is germane to the discussion we are trying to have right now about whether you can actually prevent disability or prevent, in fact, impairment, which is somewhat different from just health-related quality of life

jam

and function.

So maybe I can just show this because I think it might be helpful to this discussion. It is just an overhead about the claim, the suggestion.

DR. PETRI: I don't mind if you want to show it now.

[Slide.]

DR. STRAND: I chose to fudge and call it longevity claim rather than a specific, but simply that it would recognize, in the context of what is important to patients, the long-term duration of the disease, 20 years or so. Obviously, the trial designs that we have right now could be described in labeling, could be shorter term, and either, as I said, there could be a continuation of the trial or there could be observational trial for effectiveness to get a things beyond, say, two years in terms of effects of treatment.

The next one.

[Slide.]

DR. STRAND: We would hope that there would be controlled evidence. You would suggest there would be at least two years in duration by signs and symptoms, and I would suggest function in health-related quality of life in hopes that, ultimately, we could get to some definition of

jam

preservation of prevention of disability.

You would think that there would be difference trials and possibly act of control trials with, of course, our current agent methotrexate.

[Slide.]

DR. STRAND: The X-rays should be done yearly during these kinds of trials, but success by the prevention of structural damage might not be necessary because, for example, the horse may already be out of the barn and there may not be an ability, really, to show change by X-ray.

The analysis would give some greater weight to the end of the trial, to the longer-term benefit than to the initial, say, six months versus the last six months or an area under the curve analysis, as you were talking about, David.

Clearly, if there were too many dropouts, there were really not enough patients achieving benefit to stay in these long-term trials, observational studies, whatever, that this type of claim would then not have any meaning.

Thank you for your time.

DR. PETRI: Thank you.

I think several of the industry responses also had specific questions about how long the clinical trials would have to be for these different tiers.

jam

I think Dr. Lovell was waiting with a comment.

DR. LOVELL: Speaking to Dr. Felson's inability to Nos. 4 and 5, perhaps if it could be redefined as -- No. 4 is suppression of ongoing inflammation in a very effective fashion, and No. 5 is cure. I have no problem at all if in claim No. 5, you take very early onset patients and if you have a product -- perhaps it is overly optimistic -- that does, in fact, hit the home run and can dramatically change the outcome of this disease, then people should have an area where they could come in with a claim like that. That is what we really are all hoping for. So that would be a drug in whom you could cause complete amelioration of multiple aspects of disease with short-term application and really change the disease in the long-term way.

So I see it as a difference between studies in which you would combine enough perhaps existent therapies with enough rigor that you could suppress existent ongoing disease in a very effective fashion versus a claim for that drug that we are all hoping for which perhaps hit the home run.

DR. PETRI: Dr. Simon?

DR. SIMON: I just wanted to ask to make sure that I am not missing something here. You alluded and David alluded to the possibility of predicting who might or who

jam

might not respond, and there were two levels to that, that I am a little confused about.

One was, clearly, people who have early disease without a lot of destruction might go into remission easier because part of remission, as we created it, is somehow lack of pain associated with what has already taken place that may not be able to be repaired, but the second one is that you suggested perhaps by some markers that you may actually be able to predict who may or may not respond?

DR. PETRI: I was suggesting that all patients in these clinical trials be characterized by their risk modifiers.

DR. SIMON: But not that there are risk modifiers that may actually predict clinical response to drugs we yet don't even know about.

DR. PETRI: Correct. We don't have any crystal balls on this committee.

DR. LOVELL: I have another comment. As an alternative to looking at one or more of these claims as being subsidiaries to the signs and symptoms, look at them as independent claims, but allow companies to design trials that could address the needs of multiple claims areas in the same trial, thereby not requiring them to do a trial specifically for each claim, which I think would be

jam

inefficient.

DR. PETRI: Dr. Lovell, I don't think these can be independent. I think Dr. Felson was pointing out that there is a tier, or I was pointing out in the opposite direction that there are subsidiaries.

I don't see how you can have functional improvement in a short clinical trial without already meaning signs and symptoms. How can you have a remission if you haven't met signs and symptoms?

DR. SIMON: A remission of what? That is my problem.

DR. PETRI: A remission of visible inflammatory changes, pain globals.

DR. SIMON: There is a problem here. One is the disease which is rheumatoid arthritis, if that is the disease we are talking about, which is what we know as a biology driven by pannus formation and certain immunologic factors taking place, and one is perhaps pain and inflammation secondary to destructive disease that has already happened and already taken place.

DR. JOHNSON: It is X-ray arrest, also, though.

DR. SIMON: I understand that, but one could envision that you could actually cure rheumatoid arthritis. I realize that is not yet possible, but one could envision

jam

you could do that. The problem is you may not be able to change pain and inflammation because of the damage that has already taken place, not just progression because progression may not be related to rheumatoid arthritis once damage has taken place. You may still get X-ray progression, and yet, you have still cured the rheumatoid arthritis.

I really think it is critical to use both criteria, i.e., you remit the disease, rheumatoid arthritis, and then what do you do next from the point of view of where the patient is in the continuum of that disease? Until we do that, we are not going to be able to evaluate therapeutic implication of intervention because, in fact, we are likely going to get drugs that are going to do all of those various different stages, and this document as it stands actually does what we have done before without really doing that.

DR. PETRI: Dr. Tilley has been waiting with a question.

DR. TILLEY: I guess I am seeing the same kind of confusion from a statistical point of view because it seems to me that the signs and symptoms, as David pointed out, and the functionability are essentially intertwined so greatly that you don't have separate categories when you go from (a) to (b).

jam

On the other hand, one of the speakers pointed out the fact that health-related quality of life may be slightly separate from the functional ability, depending on the patient's perception and how they were before and how they are now.

So I think that there needs to be some work done on clarifying the (a) and (b) relationship and moving out things that are overlapped and looking at whether you really want to accept health-related quality of life as an outcome separately from the functional because linking them together doesn't make it very clear.

Secondly, this issue of structure, I think that sentence early on that was noted by one of the other speakers about the fact that you cannot have structural claims without having first shown signs and symptoms, I agree with Dr. Simon that it doesn't make a lot of sense.

If you could freeze a person in time, the way we were talking about it here, and they don't get any better at all, but they never get worse, I think that might be a valuable thing to do.

I don't think you should preclude companies from pursuing claims where they might be able to limit structural damage only.

So I think we have to really go back to this (a),

jam

(b), (c) and try to get better separation and determine what can be allowed as separate claims.

DR. WOODCOCK: This is all extremely helpful, I think.

We put "usually" in there, but obviously, that sends too strong a message. We did not intent to imply that agents that were primarily directed, say, at cartilage preservation or something like that should have to get a claim of signs and symptoms treatment first or that had some underlying biological hypothesis that wouldn't get to the inflammatory component which somehow protects the bone destruction component. So we are all in agreement. I think everyone is in agreement on that.

It is implausible, but possible, I suppose, that something that was intended to treat signs and symptoms would fail in doing that and, yet, could still arrest progression of structural damage.

I think we were trying to say -- and we can improve our language, depending on the advice of the committee -- that there would have to be some explanation, then.

When you observe a finding you didn't expect, you usually need some kind of explanation or verification of that or something like that because it could be by chance,

jam

for example.

With regard to what Michelle raised in the very beginning -- and I think as we go through, we are going to talk about the specific claims. We can talk more about some of these relationships of one claim to another. -- we had conceived that most products would be going after -- most sponsors would be looking for a signs and symptoms claim first because that is the most modest achievement as far as time span and magnitude of treatment effect required. It is the smallest treatment effect in this whole list, I think.

We had proposed or we had conceived in this document that getting a quality of life claim should not require many additional trials, and perhaps that is what you mean by subsidiary. There are many ways to do that.

The committee could advise us that they believed that the quality of life claim was so subsidiary that it should be a secondary endpoint and that it could just be mentioned in the label. It wouldn't be granted as a major claim, if you see what I mean.

We had felt that it is important maybe to elevate the assessment of quality of life because that is a major impact in rheumatoid arthritis on people's quality of life. We are all worried about it because we don't have a lot of experience in approving drugs for any drug for specifically

jam

quality of life claims. That is a new area that we are getting into.

So right now in the document, it more or less contemplates that people would go for the signs and symptoms claim, and in the same trials or perhaps one additional trial could study this quality of life and functional ability and then get that claim put right in there that this drug impacts these measures.

So that is how we conceived of it.

DR. PETRI: I just wanted to point out the danger of things like SF-36. An antidepressant might win on the SF\_36.

Dr. Chambers?

DR. CHAMBERS: I wanted to take a minute and try and differentiate between 4 and 5, at least what was in my mind as I read it, and it may not be sufficiently clear and we may need to work on that.

If you had a medication that was a cure, you took a pill and you were cured of the rest of your life, that should be recognized, and that was the purpose of 5.

Recognizing that every medication has some side effect, some downside to it, there should be some differentiation where you still had to take the medication for the rest of your life, although you may have gotten that

jam

complete response.

At least in my mind, those two were different and needed to be separated.

DR. TILLEY: I'm sorry. Are you talking about (d) and (e)? I guess we are having trouble finding 4 and 5.

DR. CHAMBERS: I'm sorry. What is (d) and (e)?

DR. TILLEY: It is on pages 5 and 6?

DR. CHAMBERS: Correct. Complete clinical response versus remission.

DR. TILLEY: Okay.

DR. PETRI: Essentially, the differentiation was whether you were on or off medication. It had to do with whether you were on or off not all drugs, but the particular drug that was being studied. I was running off of Janet's sheet.

DR. JOHNSON: But Lee's point about the fact that if you have got a drug that is curative, it is still not going to be able to attain that claim in somebody with badly deformed longstanding disease. So that is a shortcoming of the system so far.

The other one quick point, David, was when somebody was wondering about sort of gaming the system. We haven't figured a way to get around that. I mean, sure, you could load your trials with mild early patients and get

jam

substantial remission responses and perhaps make it easier to achieve that claim with those type of patients. I think that is probably the case. In fact, that was the drive to get the major clinical response in there, also.

On the other hand, we are hoping to at least be able to deal with that by describing the trials in the label.

DR. LIANG: But I think you can do that. All you have to do is get rid of that end symptoms part. I mean, I think all we are talking about is you can ameliorate the signs. A person can still be hurting from structural damage, but at least if you had some noninvasive or invasive way of saying the pannus is dead --

DR. JOHNSON: We had a long debate about that, and we couldn't get any agreement at the meeting last summer about how you would define a remission equivalent for patients with established disease.

DR. LIANG: But I am giving you the solution. I am saying get rid of "and symptoms." I can work on the signs for you. We can understand it. We can detect a juicy joint clinically, and the ones that we can't, there are newer ways of looking at pannus.

DR. PETRI: I would like to do some of these comments in order. Dr. Fernandez-Madrid has been waiting

jam

for some time.

DR. FERNANDEZ-MADRID: Well, I think I also wanted to talk about 4 and 5. I think I will agree with Dave that it is very difficult to differentiate 4 and 5.

For instance, a complete clinical response on medication could possibly be achieved during the treatment with methotrexate, but we know that as soon as the drug is continued, this patient will have symptoms and activity shortly after that. So it is not really a complete clinical response on medication because of the very short period of time.

When we use gold extensively, we could induce a complete response, and I don't call it remission because these patients eventually had activity, maybe two years, three years, four years. I have followed some patients after 10 or 15 years, and they had an exacerbation of rheumatoid arthritis after so many years. I mean, those cases, who can devise an instrument to say that these patients were cured? It is very difficult to say that.

So, in most of the cases, the few remissions that we can produce at the present time, after a while these patients clear up again. I haven't seen cured patients.

So I cannot really differentiate 4 and 5.

DR. PETRI: Dr. Simon, first.

jam

DR. SIMON: Two things about that. One is to Dr. Madrid and one is to Dr. Liang.

The first to Dr. Liang, as you may remember from a large literature, there are people that can't distinguish active rheumatoid arthritis just by clinical signs because there are subsets of patients that have very dry pannus. You biopsy them. You don't get a lot of inflammatory tissue. They are just as destructive, and some people have argued that they are more destructive based on the imperative of the fiberblast. That is number one.

Number two, I actually am very concerned that we are setting up a document that doesn't recognize that perhaps some day we may find a cure and that that cure may be real, and that because we can't measure it today, that is just our problem; that in fact, there really will be a pill some day that may make somebody better, and it may be measurably better and it may be gone as a disease. Therefore, it is not unreasonable to reward that company that devised that or the sponsor that devised that by saying that they have a major clinical remission without drug as opposed to with drug.

I do think that that is a difference because if somebody takes a pill and then is all better forever based on that disease, that is a different pill than somebody who

jam

has to stay on methotrexate or Y drug to be able to maintain that clinical response, and that clinical response may be quite significant and impressive, but not the same.

DR. FERNANDEZ-MADRID: Can I answer that?

DR. PETRI: Yes.

DR. FERNANDEZ-MADRID: I think I would agree with you. I cannot deny that such a magic pill can cure eventually rheumatoid arthritis, but we are working on a document that should be useful for industry to devise trials to prove such a claim, and I think at this moment, I don't see the way to do it.

DR. PETRI: The next comment, by Dr. Schwieterman.

DR. SCHWIETERMAN: I was going to make two points. Number one is the very same one that Dr. Simon just made that we may be on the verge of new therapies where we want this useful distinction if people think it is useful, but secondly, we are going to have a discussion of this very issue later on. I am a little bit worried about time in this.

DR. PETRI: If the panel agrees, I think there is a consensus to move on, and I think the next part is Dr. Kent Johnson and -- yes.

DR. LIANG: Actually, could you summarize what we have just said?

jam

[Laughter.]

DR. PETRI: I didn't know this was part of my job description.

We discussed the fact that many of these individual claims are interrelated; that either they are subsidiaries of signs and symptoms or there is a tiered effect.

I think we agreed with Dr. Strand that there is also a longevity aspect to some of these claims that is going to effect the length of the clinical trials necessary to substantiate them.

DR. LIANG: Well, then, before we wrap it up, may I suggest that I think that all of us who take care of patients realize this is a marriage for life, pretty much, minus 10, 15 years, and that it doesn't make sense to sort of say the trial should be this length.

I mean, what we really want, I think, is good data that helps us during the whole marriage with the patient. So I am of the mind that in chronic disease, we should just change the whole model and think about giving recognition strokes to companies that make a commitment to doing 5-year, 10-year, and that every time they passed a milestone, whether it is 2 years or 5 years, they would get a little recognition for that, you know, paper star, silver star. I

jam

think that is the kind of information that would help both the physicians and patients, and that we shouldn't just lock it in.

I think we can say something about a minimum time, but I think that what we should do is really open it up and say we want the best data you have, whether it is observational or control, but we like to have some data at every milestone in the patient's life.

DR. ABRAMSON: I would just second that and make the analogy that our diseases are like neoplastic diseases, and there is a model where you induce a remission and you have a remission for 5 years and 10 years, and at some point, you begin to understand that your intervention effected a cure.

The model from cancer is that you can't make that until 5 or 10 years have passed, and I think we increasingly have to think about rheumatoid arthritis in the same way as we develop the notion of complete remission or cure.

DR. PETRI: I wanted to assure the committee that we are not finished with this discussion, but that the FDA had additional presentations that were germane to it.

DR. FELSON: This is perhaps the central organizational point. The claims structure here sort of drives all subsidiary -- we are using the word "subsidiary"

jam

a lot -- decisions, and I think this is an important discussion.

DR. PETRI: We agree, but perhaps the presentations are going to help us focus that discussion.

DR. WOODCOCK: All right. I am going to talk about the signs and symptoms claim first, but I would first like to respond to Dr. Liang.

I think the way we give gold stars at the FDA is that you can actually put in your label of claim of durability. So, as we discuss each of these -- I mean, we aren't there yet. We don't have agents that have really durable responses of any kind that I know of. Perhaps we are close, but that would be something we could discuss. It is like the cancer model, durable response for so long. That isn't a minimum.

We need right now, as you said, to understand the minimum trial length that would be allowed.

DR. SIMON: Is that time or is there some other quality for durable?

DR. WOODCOCK: That is something we will have to discuss.

DR. SIMON: Okay. That is not predetermined.

DR. LIANG: What I am suggesting is actually a little different than what you are telling me.

jam

DR. WOODCOCK: Oh, okay.

DR. LIANG: I think we should really encourage the industry to provide data, and we should recognize it. So I don't think it is a matter of telling them that if you jump over this hoop you can get this accolade. I think that basically you should give it the accolade that whenever there is data that is systematically collected, that is helpful to people who are following people chronically.

So it is not as if this is a one-time star. It is the idea that this is a drug that has met the test for providing data over a long period of time, and I don't think it need be the fact that the person was able to stay on it for three years.

I mean, we all know that none of these are permanent successes, at least in most.

### **Signs and Symptoms**

[Slide.]

DR. WOODCOCK: We are going to discuss each of these claims in order now, assuming that there is some agreement that we should discuss each of these claims, it sounds like from the earlier discussion.

The signs and symptoms claim is really about symptomatic benefit. This claim is similar to the traditional claim that has been given for drug products for

jam

the treatment of RA.

There are some changes, however. First of all, for many of these trials, we are proposing in signs and symptoms that the trials utilize the signs and symptom observation from each time point in the trial using some kind of repeated measures.

Our reasoning for proposing this is that symptoms and signs are something that a patient is experiencing in each time point; therefore, giving a lot of weight to a single observation at the end of a six-month trial without adding in the information of what happened to the symptoms of the patient in the preceding six months of the trial doesn't seem logical because this is an experiential type of claim.

In addition, we have a lot of questions on this point. We are recommending that the trial duration be at least 12 weeks for drugs, 6 months for biologicals because of the concern about the duration of response to biological, but we have some major questions about how long this should be. We would really seek the advice of the committee on this.

The next one.

[Slide.]

DR. WOODCOCK: Now, as far as the outcome measures

jam

that might be used in these trials, there are two types of outcome measures that would be acceptable or that are recommended.

The first, which are new, were in FDA guidances at least, are use of composite endpoints. Composite endpoints have less intuitive validity to people, but obviously, they sum up a number of dimensions of signs and symptoms to a single measure.

Examples of composite endpoints that we think would be acceptable in these trials include the Paulis criteria or the ACR definition of improvement of a patient.

In addition to using a composite endpoint as the primary endpoint for a trial, sponsors could use signs and sets, sets of signs and symptoms, and these are collections of items that are observed on the patient and then some kind of statistical analysis plan is made on how to evaluate whether the trial succeeded or not.

The ACR core set, we think would be acceptable use, or the traditional four we are calling, but the outcome measures that have been frequently used in RA trials for regulatory approval, and these are the pain and swelling joint counts and the patient and doctor global assessments.

The way these have been used in trials is usually that one has had to have statistical significance in three

jam

out of these four measures, for example. That is one example of how these sets of signs and symptoms could be set up.

[Slide.]

DR. WOODCOCK: Now, just to make this a little bit more real, I am going to give an example of a trial design that could be used. You could take patients who are symptomatic, but obviously not too symptomatic, but some symptoms on NSAIDs or perhaps some additional background therapy. They could be randomized to at a placebo or the treatment, the test drug to the regimen, and followed monthly for six months with the ACR core set.

They could be scored success or failure by the ACR criteria at each time point that they were observed, and all of this would have to be agreed upon as far as when the patients would be observed.

The dropouts would get a failure score at each time after they dropped out, and then you would compare success in each arm based on a pre-agreed-upon statistical criterion for comparing the two arms.

Now, I think the committee can recognize that the score, the comparisons of the scores or the success rates in these two arms of this trial will be less intuitively obvious as to what that means clinically compared to looking

jam

at people's joint counts and comparing two mean joint counts, and that is something I think we have to accept.

There will be an ability to look behind this measure and see what happened to each of the different measures and so on, but that wouldn't be the primary statistical method of determining that a treatment was superior to placebo.

The next one.

[Slide.]

DR. WOODCOCK: Now, the questions we have for the committee, the first one, for the non-NSAID type of RA treatments, we would like to know what should be the minimum duration of trials for this signs and symptoms claim, what would be a reasonable minimum duration.

What if a sponsor decided to use only one sign or symptom as the primary endpoint, collect all the other signs and symptoms and so forth, but do the statistical analysis, primary statistical analysis on only one sign or symptom? Would that support a signs and symptoms claim if the secondary endpoints were consistent? Would they also have to be statistically superior? What would the committee think about that?

In the document, we propose that a claim of superiority to another drug should be supported by two

jam

trials, this comparative claim, and we would like to know your opinion on that.

So those are our questions.

DR. PETRI: Janet, would you prefer that we address those questions now?

DR. WOODCOCK: I don't care.

DR. PETRI: Will that be easier or would it be easier to hear the other presentations first?

DR. WOODCOCK: It would be easier to address these questions now.

DR. PETRI: I know the panel wanted to have open discussion before we actually answer these questions.

Let me start with Dr. Abramson.

DR. ABRAMSON: The question of duration, I am curious why we discriminate or differentiate for the biologicals. It seems to me that, increasingly, the drugs we are going to come forward with will have similar activity even if they are chemical. If we are inhibiting IL1 or TNF with medication, why is that different from using a biological?

So my question, globally, is why differentiate, and my sense is that probably the longer term is probably appropriate for all of these new kinds of immune modulators that are coming out.

jam

DR. PETRI: In fact, we saw data yesterday that for cyclosporine, it might take eight weeks for onset, and there was still improvement being shown as 16 weeks. That three-month time period might actually miss effective drugs. So I wholeheartedly agree with that.

Dr. Siegel?

DR. SIEGEL: Let me just clarify why it is that we suggested six months. There is a particular situation with biologics, since many of these are proteins, some of them of foreign origin, they can elicit antibodies. Sometimes those antibodies don't reach a high enough level to attenuate the effect of the drug until after three months and you can lose effectiveness. That was the reason for suggesting six months instead of three months.

Of course, none of these recommendations would preclude using a longer time frame if that were appropriate for the particular drug in question.

DR. PETRI: Dr. Lovell?

DR. LOVELL: Well, can we make a distinction? I see your point about needing to perhaps have longer trials with biologics to look for this secondary phenomena, but are we requiring that the duration of benefit be six months versus three months?

I mean, you can make a distinction between a

jam

biologic. You give it for one or two doses and you have three months of benefit. It can be just as prudent or acceptable to a patient having to take NSAIDs for three months and get benefit for three months.

I can see the point about having to follow along for the development of antibodies, but perhaps you could do the duration of the trial for six months, but not make the duration of a clinical benefit be longer for biologics.

DR. PETRI: I think I can reassure you that the dosing issues are not going to be pertinent to these claims. Those will be separate for each drugs.

Let me ask Dr. Chambers.

DR. LOVELL: I am not talking about dose. I am talking about duration of effect versus duration of kind of observations of patients in the trial.

DR. SCHWIETERMAN: That is a very useful clarification, actually, because patient benefit is patient benefit, irrespective of the type of agent they got.

Our major concern with the biologicals is the one that Dr. Siegel described. So you are quite, in fact, right. If there were to be a three-month, I would seek guidance from the committee on this point, but our perspective is if there were to be three months worth of benefit and if the committee felt that that was a sufficient

jam

length of time that we would merely require, then, follow up of those patients from the three months to assure ourselves that that benefit was durable and that we had adequately characterized that.

DR. LOVELL: My other question is about the composite criteria and the symptom courses. Have those been validated for NSAIDs? My thought was that they had been developed utilizing trial from second-line agents, and so we don't know how well those measures are going to work for NSAID-type drugs.

DR. PETRI: Let me ask Dr. Felson to answer that particular question.

DR. FELSON: Dan is absolutely right. They were validated for use with second-line drugs. They were developed -- I won't have a chance to review this, but all of the validity issues, including sensitivity to change of the individual measures in the ACR core set, were developed with a substantial database also of nonsteroidal trials.

So there is no reason to think, I guess, that they wouldn't work equally well with nonsteroidal trials because the discriminate validity, the necessity of change for NSAIDs is very good.

By and large, yes, these measures should all work very well in nonsteroidals.

jam

DR. SCHWIETERMAN: Let me just comment because that allows me, now that I have the floor, to mention something that Janet brought up which has to do with the FDA's traditional approach versus these composite measures. I think that is worth a mention.

One of the question is, is it sufficient for a company to try to get a claim for being efficacious for rheumatoid arthritis if they choose one of the four primary FDA outcomes as their measure of efficacy which might be physician global assessment or patient global assessment.

I would think that that would, in my view, not be adequate. The reason for that is many things affect patient global assessment and physician global assessment, and they aren't necessarily reflective of all of the complexity of improvement in rheumatoid arthritis which involves the improvement in swelling of joints and tenderness of joint and disability, all of those things we tried to incorporate into the core set.

That is why the core set works is because it samples broadly from the domains of the activity of rheumatoid arthritis.

If you said, well, the company is coming in with a swollen joint claim, our data suggested that rheumatologist and most people in the community felt that of all of the

jam

single measures that one might use to evaluate improvement in rheumatoid arthritis, that was the one that was most comfortable, I think, for everyone, tender joint count also being very comfortable. That is why you will notice the ACR improvement criteria requiring improvement in tender and swollen joint count because the rheumatologist felt that those were so important as measures of RA improvement.

So I think if you said a company will come in with a claim based on a percent improvement in tender and swollen joint count, forget the rest of the core set and forget the rest of the improvement criteria, I would say personally, I wouldn't find that to be that problematic.

I think if they came in with global improvement efficacy measures, that would be troublesome.

I should tell you that I don't think a company would be well served to do that because global improvements work better. They are more efficient. They give you more power than using swollen joint count, which isn't a very sensitive to change measure. So it isn't in a company's best interest to do that, but if they wanted to do that, I think they should be allowed to do that because I think it works okay.

The other problem with what you said, Janet, at least in theory, is what do you mean by commensurate

jam

improvements in the other outcome measures and when does that negate somebody's significant improvement in their swollen joint count. What do you do? Have the committee sit around and say, well, the improvement in these other measures wasn't quite as much as the improvement in swollen joint count? Is that commensurate? What is commensurate?

Then we sort of argue about what the term "commensurate" means, and I think it would be helpful to have predefined thresholds. To be honest, I think apposite measures work better for a lot of these reasons. I think they work better also because they focus on individuals with disease, patients, and I think that is much more clinically relevant than focussing on some of the mean changes in some of these individual measures. I think that is passe now.

DR. PETRI: Dr. Schwieterman?

DR. SCHWIETERMAN: Thank you. That was very useful commentary.

Let me just put this slightly in a different perspective. Because we have such useful sensitive measures now, thanks to the work of many people in the international community, we are able to use these core set criteria with, I think, a great deal of confidence.

The problem is -- and it is really not a problem -- it is a welcome problem -- is that they are so sensitive

jam

in many respects that if you have something that works really well, you don't want to use something like this because then it is obviously, as you stated, nondiscriminatory.

So perhaps I got the answer from you, but you might want to add to this. What if you have something that works really well? What are appropriate endpoints for the trial in that respect?

DR. FELSON: Well, if you have something that works really well and you are comparing it to placebo which doesn't work very well, then the appropriate endpoint is the ACR improvement criteria because the difference is going to be tremendous, and in fact, in some of the biologics being developed where there are placebo-controlled trials that are just emerging, that, in fact, is the case.

The p values are 10 to the minus something. I mean, it is very impressive discriminating ability, and I think the ACR improvement criteria work extremely well.

I think you are anticipating something we are going to talk about in a minute and that I mentioned already which is, believe it or not, since the ACR improvement criteria, I think there has been some improvement in the efficacy parameters in some of the treatments that are being tested in trials, and because of that, we may need to adopt

jam

another bar that is higher up.

Now, you will see the data in a minute that suggests that if we use that other bar, a 50 percent bar or a 70 percent bar, solely, we wouldn't have much power. It is a bad idea as a drug development idea right now because there aren't enough patients who meet that threshold to give it power.

There are still enough patients who meet the ACR criteria. So, if you have a biologic where 70 or 80 percent of patients treated meeting the ACR criteria of 14 percent or 10 percent, or 10 percent of placebo-treated patients meet those criteria, that gives you a tremendous difference, and it is easily detectable. So I think you are right. The stuff we have all developed has worked.

The good news is now things are better, and perhaps we need to create another subsidiary bar.

DR. PETRI: Dr. Miller?

DR. [CLINTON] MILLER: I want to disagree with Dr. Felson. I think composite indices came about because people did not understand multivariate analysis, and they, therefore, tried to collapse all of those indices into a single index and proceeded to refine that to the point that it was workable, but it is not the answer, and I would hate to see this document guide the future research towards a

jam

single index or composite.

What I would prefer to see is, as you said earlier, recognize that these outcomes have a multidimensional structure, recognize that there could be a tiering of what is most important second, third, and forth, and then construct that decision space with very tight limits on the most important, maybe one standard deviation of the mean, and then on the next set of modest importance, make that decision confidence limit two or three standard deviations and so on, and let your decision space have this different structure in one standard division and two in another and three in another, and continue to try to keep in your model, and your understanding of the biological events, the fact that you are looking at a spectrum of problems simultaneously to push us farther into composite.

Now, I understand that they are there. I understand that you have to have them, et cetera. I just think that would be a mistake not to leave the door open and encourage our sponsors to do that.

DR. WOODCOCK: I would like to say I think both of these points are extremely interesting, and if we developed composite endpoints, in part, because the agents had such a small treatment effect that we needed to enhance the power of our observations, also, the core measures that we use

jam

requiring statistical significance in three out of the four measures was highly unsatisfactory for reasons that are technical reasons that I don't want to go into here, but you are alluding to, I think.

We don't know yet. We hope -- there is a lot of hope -- that we may be on the threshold of having agents that have larger treatment effects in RA, and if so, I think your point is extremely well take in your point as well. We may not have the right sort of intermediate.

We have the gold standard over here, remission and so forth, and we have this composite which can detect very small treatment effects, and we don't have much in the intermediate range here, and I think if some of these actually come to pass, we may have to revisit this issue or perhaps the committee would advise us to put it in now somehow.

DR. PETRI: Dr. Pucino next.

DR. PUCINO: One other issue, if you are looking at the duration, the question is whether you have an active control or a placebo control arm, and it may not pick out the differences with an active control for eight-plus months.

DR. PETRI: Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I think I wanted to talk

jam

about the previous question.

I think in favor of Dr. Miller's proposal is the nature of the patients that are becoming available for these trials. I think the model patient that was proposed is a disappearing patient for the trials; that is, the patient with rheumatoid arthritis treated with nonsteroidals alone doesn't really reach the trials. These patients have already been treated with some type of second-line drug.

So, increasingly, we are to discriminate in a population that is complex, that has been already treated, and I think this tends to favor Dr. Miller's approach.

DR. PETRI: Dr. Tilley, and then, let's address the questions.

DR. TILLEY: I just did want to comment that later in the document, there is a door left open for multivariate analyses on page 30 when they are talking about statistical approaches. So I think the door hasn't been shut, but as we saw this morning, unless we are very clear that these are options and that there are other options available, people will be led to think they have to use the core criteria.

DR. PETRI: I think that can be handled through the wording --

DR. TILLEY: Right, exactly.

DR. PETRI: -- that the ACR 20 is one suggested

jam

approach, but that there are many other acceptable approaches.

DR. TILLEY: Right, and the same thing when you talked about the measurement tools, just being more clear, even though you said EG, be more clear that these aren't specified tools that you have to use.

My other comment related to the dropout issue which is sort of a theme that comes and goes throughout the document and was mentioned right now.

DR. PETRI: Actually, I am going to ask you to hold the dropout comment.

DR. TILLEY: Sure.

DR. PETRI: I think that is going to be discussed next.

DR. TILLEY: That will be fine.

DR. PETRI: I would like us to address the three questions that Dr. Woodcock gave us. The first is the minimal duration of clinical trials to obtain the signs and symptoms claim.

We had brought up our concern that the minimum duration should be six months. Is there a discussion before we bring this to a vote?

[No response.]

DR. PETRI: Then the question we are voting on is

jam

whether there is agreement that the minimal duration should be six months. May I see a show of hands for aye?

[Show of hands.]

DR. PETRI: Now a show of hands for nay?

[Show of hands.]

DR. PETRI: So there were four nays.

Can I ask the nays what their suggested duration would have been? I have some power as the chairperson.

DR. LOVELL: I think it is post-facto, but the reality is people who have NSAIDs they want to develop for market are -- given the time to respond for a particular medication, it may be that a patient will show improvement in two to three weeks based on the time course or response to that medication, and it is not necessary or fair to require them to do a trial where patients have to stay on medication for six months if the sponsor is comfortable that the vast majority of patients are going to demonstrate the response they are going to get in a three-month trial.

DR. PETRI: There are some other issues in terms of cellular design with this problem of regression to the mean in the first couple of weeks. I think those are complicated issues.

David, did you want to mention something as well?

DR. FELSON: I don't think there should be any

jam

duration of trial that is mandated here. I think regression to the mean can be dealt with by having a placebo group, and I think a company would be silly to plan a trial that is too short to detect maximal efficacy of its treatment.

That is their problem. I think if they can demonstrate efficacy of treatment, it shouldn't matter how long a trial is.

DR. PETRI: It is one way to achieve consensus to say that a suggested duration is six months, but that there is flexibility? No?

Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I think the reason to propose a six-month trial would be that most of the drugs that we are considering don't work very fast. Some of them may take three, six months or more to work, and I would agree not to specify in any, but if we specify three, then we are directing the industry to do something that may not really yield good results.

So I would be in favor of the defeated motion.

DR. PETRI: I think the important thing was that the nays explain their reasoning to the agency. I think that is the important message.

DR. LOVELL: If we are talking about six months as a minimum duration of a trial, then I think I agree with all

jam

the comments. The sponsors here are smart enough to know about the biologic activity of their drug and they are not going to shoot themselves in the foot by coming up with a three-month trial for cyclosporine, for example.

So I think we ought to get out of the business of delegating to people what the duration of the trial is and allow them to define based on the biological activity of the drug the minimum duration of their study.

DR. PETRI: Dr. Chambers?

DR. CHAMBERS: Can I just remind everybody that this is not just efficacy, this is a risk benefit ratio? Some of the duration here is expected to get out some of the risks while you are still in the control manner so you can evaluate the benefits in the same manner as you are doing risks. One week, I would question, or one day, even if you could show benefit.

DR. PETRI: Yes. A one-week cure if there is death at two months is not worth it.

Now, the next question we are going to discuss is whether one of the ACR core could stand along, and Dr. Felson had mentioned that the only one that the community of rheumatologist feels can stand alone is the swollen joint count.

So I am going to phrase the question is that term.

jam

Is the committee willing to have the swollen joint count stand alone as the signs and symptoms claim, swollen joint count alone?

Lee, you are shaking your head. Do you want to rephrase that question?

DR. SIMON: Well, I just don't think that that is exactly what David said. I would never suppose that I could, put words into David's mouth, but it seems to me I listened and heard that he was suggesting that there were some people in the rheumatology community that felt more comfortable that what was more reflective of disease as far as signs and symptoms go was the tenderness and swelling.

I would argue, though, that it does not stand alone; that the reason of composite index in that arena was created was because it can't stand alone. Although we feel more comfortable with its measurement, it is not sensitive enough. It is complex measurement. It varies by center, unfortunately, and therefore, multi-center trials may be difficult.

So I would argue you can't have that.

DR. PETRI: What you are telling us is that you are going to vote no, which is fine.

DR. SIMON: Yes, but I also said that I don't think that is what David said.

jam

DR. PETRI: I think our concern was that there are other things in the ACR core that definitely should not stand alone.

I mean, I made the point that an antidepressant might win. I think a narcotic might win. These have nothing to do with the pathophysiology of rheumatoid arthritis.

So Lee still gets to vote.

DR. JOHNSON: Thank you.

DR. PETRI: The question is can swollen joint counts stand alone to win on the signs and symptoms claim. May I see a how of hands for yes?

[No response.]

DR. PETRI: Show of hands for nay?

[Show of hands.]

DR. PETRI: I think there is very broad consensus on that.

The third question we were asked to address is for superiority, there will be one trial. Is that correct, Janet? Two trials? I'm sorry. The question is for superiority, should there be two trials.

Dr. Simon?

DR. SIMON: Could I ask whether it could be attended to at least two trials? Two doesn't give you a

jam

tie-breaker.

DR. PETRI: All right. The question is going to be phrased to determine the claim for superiority, two or more trials that win are necessary.

Any discussion before we bring that to a vote?

DR. LOVELL: I would like to ask Janet why that came up, why two trials versus one, that kind of devastatingly positive trial.

DR. WOODCOCK: Well, I think there could be exceptions. If a product is overwhelmingly superior, that might be one issue, but that usually isn't what you see, and so we are talking about a label claim where one sponsor is saying our product is better than this other product, and actually that has been our usual standard for comparative claims for drugs.

DR. PETRI: Dr. Chambers had a comment.

DR. CHAMBERS: Yes. I think we have viewed and I have certainly viewed it as a different threshold of what was enough to get approval as opposed to what was enough to downgrade somebody else's product and that we really wanted replication before we allowed another product which had been approved. So it had gone through the approval process and had been substantiated by trials to now be said it is not as good as something else, and we are looking for a higher

jam

threshold.

That was the purpose of this, but we are asking for comments.

DR. LIANG: But it seems to me, you have to put some -- if this is such a game, you are going to also have to level the playing field in terms of what you would accept as evidence. Are you assuming that these would stand the tests of a nitpicking academic person?

I mean, I could imagine the company would do a trial of 100 and then split the difference and report them in two journals or something like that.

DR. [CLINTON] MILLER: As a matter of fact, there is a mathematical problem there. If, in fact, they are exactly the same and the probability is one-half, then what is the likelihood of having two heads? I do three trials. Well, then I have got two out of three, et cetera.

If you just say two trials, did you mean out of five or did you mean out of six or did you mean out of three or out of two? How many of those trials were there?

DR. WOODCOCK: I am very sensitive to this issue, believe me.

I think we are trying to say that there should be substantial evidence to do what Wiley is saying, to actually make a positive statement in the label, in the advertising,

jam

that one product is better, has been proven to be better than another.

As far as the nitpicking academic, that is sort of what we do. We look very carefully at the quality of the data and the substantiation of each trial.

DR. LIANG: But what about his concerns? Are you going to also say you only get three shots at this, two out of three?

DR. WOODCOCK: Well, if the third trial showed superiority of the other agent, I think we would all have a great deal of trouble with that dataset.

DR. JOHNSON: I think you always have to look at the totality of the evidence.

DR. WOODCOCK: Absolutely.

DR. JOHNSON: If there is not good explanations for the trials that didn't succeed the way you had planned them, then I think you are in trouble.

DR. LIANG: Do you require a placebo in that comparison?

DR. PETRI: It was highly recommended.

DR. LIANG: In other words, so it would be two 3-arm --

DR. WOODCOCK: For superiority, where you are showing an actual superiority and that is a statistical

jam

test, then we don't think you need a placebo because you have demonstrated a difference. It is a difference trial.

In the equivalence trials, it is very desirable to have a placebo because you are not showing a difference between the two treatments.

DR. PETRI: Dr. Pucino

DR. PUCINO: Getting back to the comparisons of those trials, are they all going to use the same outcome measures? There is still no consistency with what type of outcome measures.

As Dr. Liang alluded to, what is the evidence, is the evidence consistent?

DR. PETRI: I think the consensus of the committee is that this should not be a question; that in fact, it is a statement of fact that the totality of the evidence should be in favor of the drug if there is going to be a superiority claim.

Is there a consensus?

DR. JOHNSON: But we do want some consensus as to whether one trial is enough. I mean, that is key, really, I think.

DR. PETRI: I am not sure how we can address that, one trial done at 25 different sites where all sites show superiority of the drug and all of the risk modifiers are

jam

accounted for.

DR. JOHNSON: That might fall under the exception rank that Janet alluded to, as would a home run. As to the companies, they are interested in our feedback and your feedback about one or two trials, too. I mean, it is a big hurdle to stick in another trial.

DR. PETRI: Dr. Strand wanted to comment.

DR. STRAND: All RA trials, if they are trying to show superiority by definition will be at 25 sites. So I am not sure that that kind of a thing is going to give you enough robust information.

DR. PETRI: I think one of the objections to one trial would be if it had been at just a few sites, and there might have been some systematic bias.

Dr. Abramson.

DR. ABRAMSON: Yes. I am curious, just first, how the FDA deals in other fields with superiority of drug because it seems like we can't be comfortable with one trial or two trials.

Our discussion earlier, as we try and evaluate these drugs with regard to partial response, major response, complete response, is I think how each drug is going to have to be evaluated. How we come to decide that, then drugs can be compared in the marketplace, but I think our job might be

jam

simply to say a given drug induces a partial response or a major response in X percent of the people who have been evaluated, and then, that is the best we can do, in my mind, to compare one drug versus another, given the dilemmas that we are talking about with the nature of the clinical trials one or two or five.

DR. WOODCOCK: All right. Well, maybe we will take that as your best advice, and we will decide what regulatory recommendations we make for sponsors. We could do that.

DR. PETRI: I believe we are up to the next part of the discussion now, which is a pro/con debate about how dropouts should be handled.

I believe Dr. Johnson and Dr. Siegel will be in charge of this debate.

**Pro/Con Debate re: Dropouts**

DR. JOHNSON: We are actually not going to really do a debate. We are going to do sort of a pro/con, though, and this is an aside, but I think it is an important aside.

One of the reasons we are having this meeting is to really kind of try to get some help on the nuts and bolts problems of trial design and analysis. In my opinion, this is a big problem trial designs.

So what I am going to do is present you with an

jam

absurd example, but an example that is internally logically sound, whereby both the conclusion either done by intent to treat or the conclusion as done by a completed analysis, which are the two traditional ways, both show the result to be highly statistically significant. Yet, I am going to show you that the result is wrong.

The way I am going to be able to show you the result is wrong is because I am going to start with the assumption that the drug does nothing, okay?

So this is the title of the talk, "A differential dropout," and actually, at the end of all this, I am going to just list some techniques whereby dropouts can be minimized. Two points actually, dropouts can be minimized and the importance of close follow-up of the dropouts that do happen to occur, and then Jeff is going to take it from there.

[Slide.]

DR. JOHNSON: Here is the mechanism that we worry about, that I worry about, and that I think has some credence in prior trial experience.

One, severe drug patients may drop out, so that the completers are enriched with mild patients, and mild placebo patients may drop out, so that the placebo completers are enriched with severe patients.

jam

Next slide.

[Slide.]

DR. JOHNSON: Now, why might this happen? Well, you could fancy that, for instance, a severe patient who sensed that he was on a drug that had some mild side effects, such as in the nonsteroidals -- and by the way, what this means is that there is already a little bit of unblinding that has crept in.

He may drop out if he is not doing well because he says, what the hell, why should I bother with this.

[Slide.]

DR. JOHNSON: Similarly, a placebo patient who has just mild disease may, for whatever reasons, sense he is on placebo, and he may think, well, this is not worth it, why should I bother with this trial, I am not getting any benefit because I am on placebo. So he drops out.

Next slide.]

DR. JOHNSON: What I am going to show is an example where, as I said before, a highly statistically significant result by either type of analysis occurs; yet, the conclusion is wrong.

Next slide.

[Slide.]

DR. JOHNSON: So what you assume is that the drug

jam

is inactive, and that is a handy assumption because then you can infer the correct answer, and the correct answer should be that no effect is shown unless chance alone happens to show it, but we will leave that possibility aside.

So here is the assumptions of the trials or the characteristics of this hypothetical trial, a six-month double-blind placebo-controlled, two arm, 100 patients per arm. A standard dropout rate, let's say, is 20 percent.

We use a by-patient success test, and we can just define it as somebody who completes the trial and they improve by at least 25 percent, let's say, over their baseline.

It doesn't even really matter what your baseline value is. I mean, what measure is being used in this trial? You can take any arbitrary measure, composite measure, individual measure.

[Slide.]

DR. JOHNSON: I also want to, for simplicity, just assume that rheumatoid has three discrete courses. I mean, we all know it is a variable disease. So this is a simplification of reality.

A quarter of the patients improve by 30, a half show no change, and a quarter are worsened by 30.

[Slide.]

jam

DR. JOHNSON: So we have done the trial, and you go to the end, and of the 20-percent dropouts, which as you recall are assumed to occur in both arms, there is a differential dropout phenomenon, as driven by those perceptions that I had mentioned earlier.

So, regarding the drug patients first, the 80 drug patients complete the trial, but since the 20 drug dropouts were severe and so unresponsive, none come from the 25 doing well, all right? So there are 25 completers who -- leaving 25 completers, 30 percent better. Thirty percent better is the number of successful patients. It should be 25 percent better.

[Slide.]

DR. JOHNSON: There are 25 who satisfy the criteria for being a successful patient, and if you look at the placebo arm, the 80 placebo patients also complete the trial because 20 of them drop out, but the dropouts here are mild patients and so more likely to be responders and so come from the 25 who are doing better, leaving only five completers left with the -- five patients who complete by the test of -- by patient success which is 30 percent better than baseline.

Next slide.

[Slide.]

jam

DR. JOHNSON: So, if you do an intent-to-treat analysis, which is of all randomized patients, first of all, in the case of the drug arm, you have got 25 over 100 that are successful patients, and in the placebo, you have only got five over 100. So that is nice and statistically significant.

[Slide.]

DR. JOHNSON: Likewise, regarding a completer analysis, you have got 25 over 80, 25 completers over 80. You have got 25 success over 80 completers for the drug, versus only five successful completers over 80 in total for the placebo, which is also highly statistically significant.

Next slide.

[Slide.]

DR. JOHNSON: So here we have success to a high degree of significance by both ITT and completer for an inactive agent. How can this be avoided?

Next slide.

[Slide.]

DR. JOHNSON: I think, logically, there are two categories that you could divide up techniques into, and one would be to avoid entirely or minimize, if that is the best you can do, discretionary dropouts, and number two, if you do get dropouts, and you can't really forbid dropouts -- you

jam

have to follow them to the end of what would have been their trial duration and collect some data, so that you can try to refute the assertion that the results were due to dropouts.

Next slide.

[Slide.]

DR. JOHNSON: In the first category, just some ideas about how to avoid discretionary dropouts. In trial design dimensions, that may help in this regard. You can outlaw it, preclude discretionary dropouts, but that is obviously unethical.

You could assert or feel that you design so that you had absolute certainty of 100 percent blinding, but that, too, is probably impossible, at least in most cases.

You could stir up investigator enthusiasm and hope that that translates into patient enthusiasm, and I think that has played a major role in the past, and I think it still even plays an interesting role differentially across from one country to another. I think some of the other countries don't have as much a problem with dropouts as we do.

This kind of enthusiasm pertains to both patients and physicians, and there are various ways you can reward people with things along the way or with a promise of the magic drug at the end of the trial and so on. These are

jam

valuable, but I think there are some limitations as to the degree you can deploy these.

Another idea is to increase compliance by having the whole trial be passive treatment, essentially. In other words, if you have got a wonderful induction agent that only has to be given once, then you are really not talking about ongoing medication compliance, anyway. You have just got to get them to come back for those visits to be assessed, but you need a good drug if that is the case.

[Slide.]

DR. JOHNSON: So notions that are external to the trial would be certain obvious things. If you have a great drug or if you have a drug with minimal side effects, those both would be helpful.

Have there be no other treatment options, what this does is enable you to have a lot of confidence in what the natural history is because, presumably, if you have got terrible patients who aren't going to spontaneously remit -- but this means your drug is following a strategy of a drug of last resort, which is not what companies want to do often.

You can study a progressively fatal disease, and there again, that gives you this kind of confidence, but that is not usually the case in rheumatoid arthritis.

jam

Two more slides.

[Slide.]

DR. JOHNSON: Follow, analyzing dropouts. The trial design mechanisms here, I think, are important, especially in the protocol. If you specify recall of dropouts and don't make it optional and don't make an extension study and you have to cook it up as routine care, essentially -- but obviously, there are limits as to how much off study data you could collect.

Another point, record major confounders, I think this is sort of self-evident, if you think about it, because you are going to need that information to marshal your arguments that the dropouts don't undermine your conclusion, but here again, it is incomplete data, and it is open data, too. So those caveats have to be kept in line.

Finally, you could incarcerate your patients and follow them up, but obviously, that is impractical.

It is interesting, some of the early NIH work did have more closed populations or populations which made follow-up easier. I just add as a postscript, which you probably can't read there, the comment that I don't think it does take a lot of follow-up data to enable one to marshal a convincing argument against the differential dropout contention, but it almost always take some compulsively

jam

collected follow-up data.

Thank you.

[Slide.]

DR. SIEGEL: The issue of how to handle dropouts is a major issue of discussion between the FDA and sponsors in the design of clinical trials especially in the design of Phase 3 clinical trials.

[Slide.]

DR. SIEGEL: One of the difficulties in knowing how best to handle this is that different techniques for analyzing dropouts may bias a trial result either for or against the agent in question.

[Slide.]

DR. SIEGEL: So what I would like to do is present to you a variation in clinical trial design which is intended to minimize the number of dropouts.

The problem, just to restate it, is that high dropout rates in clinical trials compromise the analysis from rheumatoid arthritis trials, especially long-term trials. Yet, on the other hand, longer-term compared to efficacy trials are desirable in order to demonstrate a durable response, and particularly to assess the effects of therapy on structural damage.

[Slide.]

jam

DR. SIEGEL: The solution that I would like to discuss this morning is the idea of incorporating subject withdrawal due to lack of efficacy into an endpoint which measures comparative response rates.

Now, a design like this is appropriate for agents with certain characteristics and possibly inappropriate for other agents. We have begun to see results of Phase 2 clinical trials in the Center for Biologics, newer agents which have particular features.

We are seeing some agents that have high-response rates, using an ACR-20 criteria of over 50 percent in some cases, a rapid response measured in weeks, mean responses in some cases of one or two weeks, long-lasting responses, and low dropout rates in the treatment arm, so that you have the unusual situation where the dropout rate in the treatment arm is much lower than the dropout rate in the control arm, the placebo arm.

What I would like to present is the case for agents with features like this. Non-responders are easily identified and may be incorporated into the endpoint.

[Slide.]

DR. SIEGEL: Now, I think the experimental hypothesis is slightly different, and I would like to call this kind of a trial a sustained response rate endpoint

jam

trial.

This is an example of how a trial like this might be set up, and of course, depending on the results of Phase 2 trials, the particular criteria that are used in the protocol may differ, but basically, the endpoint in this hypothetical trial would be to compare the proportion of patients with an ACR-20 response in six months.

In addition, after two months of therapy, subjects with no more than a, for example, 15-percent improvement on two consecutive clinic visits would be declared a non-responder, and subjects who stop therapy due to toxicity would also be declared non-responders.

These non-responders would be removed from treatment, but continue to be followed to the end of the study, and they would be classified as non-responders at the final endpoint.

[Slide.]

DR. SIEGEL: I would like to point out that in a study design like this, there would basically be one category of patients who would be declared successes. These would be subjects who stayed on the study drug for six months and had an ACR-20 response at the final six-month endpoint.

In contrast, subjects to fail in this study might

jam

fit into one of three categories. The first would be subject to stay on the study drug for six months, but do not have an ACR-20 response at the end.

The second category would be subjects who have 15 percent or less ACR response during the trial and, therefore, stop treatment due to lack of efficacy by protocol-defined criteria.

Finally, subjects with significant toxicity would be declared non-responders.

[Slide.]

DR. SIEGEL: The potential benefits of a design like this would be that you would minimize lost data. You would avoid prolonged treatment with ineffective therapy, and this gets around some of the ethical concerns with placebo-controlled trials, and in some ways, a design like this conforms more closely to clinical practice where if a patient isn't responding to therapy, they wouldn't be continued on therapy indefinitely.

I would just like to give you an example of a hypothetical drug which is effective and how a trial would be analyzed using several different ways of analyzing and accounting for dropouts.

[Slide.]

DR. SIEGEL: Suppose that for this hypothetical

jam

drug, the percent of ACR-20 responders at six months were 45 percent, compared to 20 percent on a placebo control. The remainder of the patients are non-responders, and in the drug arm, those would be 55 of whom, let's say, half drop out month two through six, or 27 out of 55 dropouts.

In the placebo arm, there would be 80 non-responders, and again, half would drop out due to a lack of efficacy during the course of the trial, namely 40 out of 80.

The completers in this case in the drug arm would be the 45 responders and half the non-responders for 73.

In the placebo arm, there would be 60 completers.

[Slide.]

DR. SIEGEL: Let's see how the analysis would come out with three different mechanisms of analysis.

Using an intent-to-treat analysis in the left side, there would be a 20-percent response rate because 20 percent of the patients had completed the trial, 20 out of 100. With drug, it would be 45 percent.

In contrast, there would be a 45 percent dropout rate among the placebo group, 27 percent in the drug group, and this differential dropout, 40 versus 27 percent, would raise questions about whether there was differential dropout, and this might compromise the analysis of the

jam

result.

If you use a completer analysis, you would get a 33-percent response rate in the placebo arm, a 62 percent in the drug arm, again, 40 versus 20 percent dropouts, and this completer analysis would inflate the rate of success in both groups. So it would distort the results.

In contrast, on the right side are the results with the sustained response analysis such as I described before. Here, you would get a 20 percent response rate in the placebo arm, 45 percent in the drug arm. There would be no dropouts. Namely, all the patients who had lack of efficacy would be protocol-declared failures. So none of them would be dropouts in the sense of being unaccounted for. In this case, you would have a complete dataset, and it would give you greater confidence in the result.

I have to point out a couple of caveats. One is that if you had a waxing and waning disease course on drug, this might not be an appropriate endpoint to use.

Secondly, there are blinding issues, and I think that it might be particularly helpful in a design like this, as well as perhaps in other study designs, to use an independent joint assessor who has no knowledge about the clinical course of that particular patient to perform the essential joint analyses for the endpoint.

jam

I think it is important to point out it is important to continue to follow the non-responders, and subjects who drop out for reasons besides toxicity or lack of efficacy must be assessed for differential response.

What this says is that there are always going to be a few patients who drop out because they moved to another area or lost to follow-up, and these patients must be assessed to make sure they don't bias the result.

[Slide.]

DR. SIEGEL: So the question that we would like to address to the committee is that does this trial design represent an effective way to address the problem of information lost due to dropouts.

DR. PETRI: I think the committee will remember that in the guidance document and in the response from industry, there was a lot of concern that a requirement of 85 percent of patients complete a trial is going to be almost impossible to meet. So this is an alternative.

Is there discussion before we bring this to a vote?

Dr. Tilley?

DR. TILLEY: Is there commonality in the two discussants? One, the main point I think they both stressed was the importance of following people who have gone off

jam

medication, and really, the key difference that I saw was that one -- our second speaker was giving us criteria for taking people off medication, and to me, that is really a separate issue. I don't think it really needs to be a part of a dropout discussion. I think the problem has been the synonymous term of dropout and off study medication. They are really two completely different things.

A dropout to me is someone who you can't find information on at the end of the trial. A person who is on or off study medication is someone you are looking at as part of the understanding of your trial and the interpretation of your results, but I think the key feature here is follow everybody to the end.

DR. WOODCOCK: That is a very, very useful distinction, and I think we should incorporate that in the document.

I think I would like to say the way the two speakers presented these two different ways of doing trials, the first method or some of the things Kent was promulgating as far as still doing intent-to-treat type of analysis will give you more of an estimate of the magnitude of the treatment effect. Now, that is a traditional thing that we usually do.

The type of trial that Dr. Siegel was discussing

jam

will give you treatment policy, the effect of your treatment policy, whether you choose this medication, how many patients will be defined as successful at the end of six months, which is very useful for clinicians, but doesn't really give as well or as intuitively. I don't think the magnitude of what you achieved in that -- if everybody follows me.

I think the problems with the dropouts are so severe that if we could make that distinction, I think it is incumbent upon us -- and the committee could give us advice about this -- any mechanism allowing alternative treatments while the patient goes off study medicine, but remains in the trial, many other mechanisms that will keep people in observation, although maybe not on treatment, the experimental treatment would be extremely desirable.

DR. PETRI: Dr. Felson first and then Dr. Lovell.

DR. FELSON: I also sort of wanted to echo a little bit of what Barbara had said in the sense that I think the solution to this -- dropouts are impossible to deal with in a valid way, and I think that the solution to this problem is to do everything one can in the design of the trial and in the FDA's approach to acceptable trial design that will lead to structural trial which minimizes dropout, and in that vein, let me suggest that now that we

jam

-- I should perhaps assume that we have effective therapies for rheumatoid arthritis, known to be efficacious, and that one of the approaches of a long-term placebo control can't go on any efficacious therapy if you are randomized to placebo unless you drop out of the trial approach -- is inherently problematic, and it is going to lead to dropout rates. It is going to lead to all of these problems that we are talking about.

So, in the context of later discussions where we talk about comparability or equivalency trials, where we talk about superiority to other drug designs, I think the FDA should strongly consider incorporating those into approval so that not so many patients need to be -- "assigned" is perhaps too euphemistic a term -- to placebo. I mean, have to be penalized by being place don placebo for a period of time long enough that they suffer, that they may develop disability, that they may even develop structural damage which could be avoided by use of efficacious therapies which are already noted.

The Tugwell Wells, et cetera, design of the cyclosporin trial that we heard about yesterday was another alternative around this, which is to test all of these new drugs as marginal therapies on top of second-line drugs so we don't penalize our patients and force them to come off of

jam

the second-line drugs that they are on.

Any number of trial designs would be acceptable here, but to be honest with you, I think the way around the dropout problem is to not force drug development with people off all second-line drugs and randomize, having half of the patients randomize to placebo. That is asking for big problems.

The CSSRD trials, 40 percent dropout, placebo dropout rates in a year period. The Gold placebo trial, a notorious trial, people had to come off their second-line drugs, got randomized to placebo for a year. Most of them couldn't tolerate it. I mean, you know, it is just not going to work. There is no amount of little encouragement we could give to patients whose disease is flaring and more miserable.

DR. WOODCOCK: I would like to make this real clear. I think it is an excellent point.

The document maybe -- perhaps the committee could give us some advice. Maybe it isn't clear enough on this. There is absolutely no intent in this document to say that these should be long-term placebo-controlled trials.

As I said at the beginning, there are three kinds of designs you could use to show simple effectiveness. One is verus placebo, but that placebo can be on top of all

jam

sorts of background therapy. The patient could be on gold methotrexate and NSAIDs, and you could add, provided you had assurance of safety -- you could add the new experimental treatment versus placebo on top of that.

You can do active control trials or you can do superiority to existing therapy trials. There is still a dropout. So let's say that is a given. That is all a given. It sounds like most people are in agreement with that.

Then the question is how do we still handle dropouts. You take people on methotrexate gold and NSAIDs and you randomize them to add placebo to their regimen or active treatment. They still aren't doing well.

What I was saying is I think we need to explicitly have mechanisms where they can be pulled off of their active treatment, remain in the trial. They may have to go on other therapies, but we can still follow them. That is the confounder problem that Kent was talking about, but it is probably better than actually losing those patients to observation altogether.

DR. PETRI: Dr. Lovell was next.

DR. LOVELL: I am a little confused as to what you do with those patients at the end of your trial, and this is a question of ignorance I think.

jam

So, on Jeff's study design, you had patients who kind of were defined as failures at two months, and you followed them for an additional four months, so that they would be around at the end of the six months.

It is clear that in terms of toxicity and safety, that is a desirable way to handle it, but from defining efficacy, because I assume they are not going to live in a vacuum for a month, so they will be trying something with their medications. So you have those patients back at six months and do their evaluation. How do you, in fact, handle the information you get from those patients?

DR. SIEGEL: While you would follow those patients, as you point out, they would be on a variety of other therapies, and they might have a successful response to those therapies. I think that that information should not be considered in the assessment of the primary endpoint; that with a trial design like this, you would use -- at least for the primary response to treatment -- it would be the criteria that I suggested.

DR. LOVELL: I guess actually a more appropriate question would be if we use it in the intent-to-treat, kind of your scenario, the alternative, the more traditional way of analyzing data.

Let's say we are able to successfully keep these

jam

patients coming, even though they are not part of the trial, and we give the information at the end of the designated study period. How would you utilize that data in analyzing the effect of this drug?

DR. PETRI: I think that was answered by Dr. Siegel that there would be a sustained response analysis.

DR. LOVELL: It is answerable in his study design, but it is not so obvious in the more traditional study design.

DR. WOODCOCK: Right. I think in that study design, they are already declared failures. So that is why you lose sort of the treatment effect information, but because they are confounders in the intention-to-treat analysis, you probably would do a with and without dropouts analysis, and that would give you some idea of their robustness of the primary analysis, which probably would not include dropouts who had gone on to other treatments.

I don't know the answer to this.

DR. TILLEY: Can I make a comment on that from the statistical point of view? I guess I am a firm believer in intention-to-treat analyses, and if my outcome was responder or non-responder, it is not as difficult, but in a more complex trial where you are looking -- for example, you are using the ACR criteria, I would measure the ACR criteria at

jam

the end. If they went on something else, that is the way it is.

You want to know was there a difference between your two groups at the end, and I would do the intent-to-treat analysis, which means it would be harder to find differences. It means that your treatment would have to be a lot better, but it is certainly a reasonable choice, but there is noise in this whole population, and there are flares. This is, as you all know, a difficult disease, and I think the most rigorous approach we can bring to it is the intent-to-treat.

DR. PETRI: Industry will have to create incentives to make that kind of an analysis work.

DR. TILLEY: Yes.

DR. PETRI: Dr. Gorelick had a comment.

DR. GORELICK: I have two points that I would like to make. Number one is that we talk about dropouts, but I think we are also including noncompliance patients in that population. One of the issues that I have had is you can take a look at two different types of examples of effective treatments that in an intent-to-treat setting might not come out appearing efficacious.

I think not in this particular area, but if you take a look at, for example, the use of condoms in HIV

jam

prevention, they work very well in use situations when you actually are using the agent. However, in a controlled clinical trial thing where you are not observing patients, they may not be used, they may not be used correctly, and therefore, you would come to the erroneous conclusion in your controlled clinical trial that the product didn't work.

Similarly, in anti-tuberculous therapy, we know that agents only work when they are used, and we have moved to a direct observation of treatment method in a large group of patients because we can't be certain about compliance, and again, if we assume that compliance was occurring, i.e., there were no dropouts and patients were taking the drugs, we would possibly come to the wrong conclusion about the product.

So it is a question I have, and assuming that all dropouts are failures and everything else, we may come to the wrong conclusion about efficacy of an agent.

The second point is that in a clinical setting -- in a clinical trial setting, we are really treating patients on a wholesale basis. We are trying to look at population effects of a drug, but in a clinical practice, we are treating patients on a retail basis. We are looking at our patient in front of us and asking does this treatment work.

I think that neither of the approaches that I have

jam

heard presented here really will help us to come to a conclusion that a drug works in a particular individual, and I haven't yet heard any trial design that helps us to come to those conclusions, and I am concerned that we are discussing some fairly far-reaching approaches to clinical trial management, and I would like some comment.

DR. PETRI: Dr. Johnson?

DR. JOHNSON: I am not going to try to answer that question. I think that is the efficacy versus effectiveness differentiation, which is in my opinion over-polarized, but I think that is a whole other debate.

I think David is correct. There is no solution to this, and if you do it even further than an intent-to-treat with less observation brought forward, if you do a worst-case scenario where the placebo dropouts get the best score subsequently and the drug dropouts get the worst score and your drug still wins, then that would refute the argument that the dropouts are a problem, but that obviously is an incredibly conservative approach.

It is true that this problem really consumes noncompliance, too. The extreme of noncompliance, you could say, is dropping out.

And even with Jeff's design, if all your placebo patients who -- or all your non-responding patients who were

jam

required to drop at two months, at six months were remitted, what would you do? I mean, that obviously would blow your analysis right out of the water.

The reason you would say that that is very unlikely is because you have confidence in the natural history of the disease saying that that is not going to happen, but if we have confidence in the natural history of disease, that is a huge conceptual leap forward, and it would allow us to do all kinds of stuff.

DR. WOODCOCK: I would like to answer Dr. Gorelick's question or make some suggestions.

Actually, I think the design that Jeff Siegel presented is a retail approach, very highly. It is a treatment policy.

If I adopt this treatment policy, how many successes will I have? It is a by-patient analysis for success that is being done, and it is a number of patients who succeeded final analysis that is being done. So it has some of the caveats that have been raised, but it is very much from the perspective, I think, of the treating clinician if I set out to treat this patient. One of the chances are that that patient will be better at the end of the day compared to if I didn't treat them at all.

DR. PETRI: Last comment, Dr. Schwieterman.

jam

DR. SCHWIETERMAN: I was just going to point out that, by no means, do we consider this to be an ideal trial design. There are a number of considerations, including late-acting effects, including waxing and waning disease courses and so forth, but again, the reason we are faced with this, in many respects, is because our prior -- from past years studies are such that sponsors believe their agents to be quite effective and are reluctant for ethical reasons and other administrative reasons to allow patients to be on an ineffective therapy for a given duration of time and are approaching us with various solutions to that problem, namely how you can protocol-define what treatment failures are so that those patients aren't necessarily given that treatment.

To the extent that you can use a differential dropout rate as an outcome measure, this approach makes sense, but I think Dr. Tilley's point was well taken that there are more rigorous methods that you can use, intent-to-treat analysis, that we would also employ with this type of design.

It is an interesting concept, and I think the discussion has been informative.

DR. PETRI: The question we were actually asked to address is near the bottom on the second page which is:

jam

Does such a trial design represent an effective way to address the problem of information loss due to dropouts? I don't think we need to bring this to a vote. I think there is a consensus of the committee that there are probably several trial designs, as discussed, that will help minimize this problem.

Now, we have one other section that we would like to cover before the break, and this is on function/quality of life, and then structure. So if we could move ahead and do those two things before the break.

Kent Johnson is going to discuss the function/quality of life.

### **Function/Quality of Life**

[Slide.]

DR. JOHNSON: I think the discussion is much better than the presentations.

This is the claim that I have been asked to talk about or that I volunteered to talk about because I couldn't get anybody else to talk about it, improvement in functional ability and health-related quality of life.

[Slide.]

DR. JOHNSON: The background involves some definition of concepts. We have talked about this already today a little bit, the notion of physical function is a

jam

little more intuitively obvious.

Claire Bombardie and other people have conceptualized this process as a continuum for pathology to impairment to functional limitation to disability to handicap.

[Slide.]

DR. JOHNSON: Disability and handicap are seen and, hence, measured and quantified in the context of individual and sometimes societal or social expectations. All five in some ways impact an individual's perceived quality of life.

I think this point is important, too. It is not obvious that an improvement in function translates into an improvement in disability or vice versa, and more importantly, that a loss of function necessarily becomes a disability.

We have to specify not just quality of life, but health-related quality of life. Quality of life, in general, entails a lot of domain, some of which I have listed up there, and health-related quality of life, as has been pointed out earlier, also has a number of domains.

I don't want to get into the nuances of the instruments, but any health-related quality of life -- could I have the next one?

jam

[Slide.]

DR. JOHNSON: -- is by necessity filtered through an individual patient and also often societally based, but I think from our point of view, it is important to try to capture the full concept, in other words, always that the disease impinges on domains that are important to the patients.

I think it is also important to recognize that what we don't capture here -- and that may be certain drug toxicities. More properly, you probably need utility measures to get a full balance of toxicity versus benefit, and as you know, traditionally, the FDA has conceptually separated assessment of toxicity in arthritis trials.

I would argue that the rationale for this kind of claim is sort of self-evident. In other words, that the health-related quality of life has faced validity and that there is no argument about that.

I hope Peter agrees with me there.

[Slide.]

DR. JOHNSON: In addition, there have been a number, as you know, of multiple international consensus conferences to try to come to grips with these concepts. So, accordingly, we have construed this claim that we are lumping together function and health-related quality of

jam

life, and in pursuing it, we are asking for the use of a validated functional measure in RA of which there are a number that already exist, and I have listed a few of them here.

[Slide.]

DR. JOHNSON: In addition, we are asking for a validated health-related quality of life measure, either in RA-specific measure or a generic measure shown sensitive to RA, and I think the one at this point in time that has been used the most in the States is the SF-36, which is a generic measure, and, hence, gives the advantage which isn't really strictly a regulatory dimension, but is useful for health policy in general to bench mark rheumatoid versus other diseases, essentially.

[Slide.]

DR. JOHNSON: Finally, as you know, we are at this point ordinarily or should also improve signs and symptoms. That is open for discussion.

The last slide.

[Slide.]

DR. JOHNSON: So we are going to ask you to address these three questions. One is this difficult question of the duration, and it has come up earlier this morning. If you are going after four claims, you may not

jam

need eight trials. I mean, there is no problem with doing one trial and pursuing more than one claim within that trial, although there statistically sometimes becomes an issue of multiple measures and primary and secondary measures and so on.

As has been pointed out, duration implies greater and greater challenge as these things get of longer duration for drug developments to consider, if it is a 12-month trial versus a 6-month trial.

Finally, can this claim stand alone or should it stand alone or should it stand -- you know, if you think that maybe it should be a 6-month claim or a 12-month claim, what if in that trial you only do -- what if in a previous trial, you have done signs and symptoms for three months? Is that adequate? And if you say that is adequate, then what if signs and symptoms don't do well in the 12-month trial. I mean, there is a lot of permutations to this, if you think about it.

So I am going to stop there and turn it back to Michelle.

DR. PETRI: I think we started this discussion this morning, but let me ask for additional comments.

Dr. Tilley?

DR. TILLEY: What was the rationale for putting

jam

the two things together, function and health-related quality of life? What was the rationale for combining the two?

DR. JOHNSON: Well, I think it was our prejudice that long-term -- I think it partially emanated from this impression that we have to try to think of longer-term outcomes in rheumatoids, and it doesn't sort of really matter if your joint counts are five or four or there if you can't walk.

So we wanted physical function, and we wanted to -- and we wanted to have it -- and long-term outcome has to reflect quality of life. That would be my answer to that question.

DR. PETRI: Just to remind the committee about the previous discussion, I mean, one point we made was that function was already part of the ACR-20. It was closely related to signs and symptoms. The second part of the discussion we had this morning was that several drugs have nothing to do with RA pathophysiology. It could easily win on an SF-36, such as a narcotic or antidepressant.

Dr. Felson?

DR. FELSON: Let me suggest that there are two questions embodied here. One is whether we should be distinguishing in claims or trials between physical function change and the broader concept of health status or quality

jam

of life, and those aren't necessarily synonymous, but let's call them synonymous for now.

The other question is one, I think, that Dr. Woodcock raised earlier, which is time. We know that this is a disease with a long-term disabling trajectory. Do we want to do something to try to encourage companies to go after that and try to prevent that trajectory?

Let me speak briefly to both of those questions. I am sure there will be other people that will, too. I can tell you, analyzing data from functional status instruments, especially AIMS which we have a lot of experience analyzing, that the sensitive measure here is physical function, that that is what changed in RA trials.

It is not clear to me that we should ask a company to evaluate whether RA affects emotional function, social role, all of those other things subsumed in quality of life.

Don't get me wrong. I think they are very important. There are important global concepts of quality of life, but I don't think that is necessarily what RA drugs should be targeted to do.

I think it is critical that we ask them to affect physical function, which is such a fundamental component of what is important in rheumatoid arthritis.

By the way, the SF-36, if you talk to John Worth,

jam

the arthritis meetings, it does well in arthritis because it happens to have questions on bodily pain and on fatigue, which are very important symptoms in rheumatoid arthritis. The emotional stuff in it isn't affected by RA. It isn't affected by RA improvement in the context of a trial.

A lot of the components of health-related quality of life don't necessarily change very much when RA is successfully treated because they are not really that relevant to RA. They are affected by a lot of things.

So the next question is what time do we think about. I think that is a very interesting question, and I think it wouldn't be a bad idea to think about putting physical disability, a target for a claim subsumed under signs and symptoms that says, look, we will give you a claim that your drug affects physical disability if you can show over two years, not six months, not even a year, but the long-term-type window that Ted Pincus and Fred Wolfe have been talking about beginning in over two years in a trial that your drug compared to nontreated patients somehow affects that trajectory. Now, that is a big hurdle. That is a very expensive long-term trial, but nonetheless, I think that is beginning to be the window that we are all interested in. This is a long-term disease, and I think we ought to ask if somebody wants that claim that they try to

jam

get out to that length of time.

DR. PETRI: Additional comments, Ms. Malone?

MS. MALONE: I agree with Dr. Felson. I think you can measure the physical functional ability, which in most cases, if it increases or becomes more positive, it will usually enhance quality of life.

The quality of life is so subjective, and to someone with a long-term disease, so much depends on their emotional makeup, their psychosocial environment, their attitude, their attitude towards life, their support system.

I don't see how we can measure it.

DR. PETRI: Additional comments?

Dr. Liang?

DR. LIANG: Well, we do.

MS. MALONE: I know we do, but you are subjective. This is the patient who is living through it.

DR. LIANG: Actually, these are mostly derived from the patient, all of the measures that we are talking about, which is more than a 45-year history in rheumatology. Actually, it is the paradigm shift that takes the patient's view into account, and they are very powerful predictors, and they have excellent psychometric properties.

I agree with David. I think that -- and we have made this point repeatedly -- with any of these wonderfully

jam

psychometrically proven instruments, there are floor and ceiling effects, and I think if you are an experienced clinician, you actually follow patients over time. You realize that you sort of come to the end of the rope in these scales very oftentimes, and the patient can still benefit.

I think that we are sort of past the point of saying this is new technology waiting for evolvment, which is the first extramural talk we had this morning. I think they are here and now. They are part of the literature, but I think we also recognize the limitations.

I still think that there are still basically self-reported measures of function, multifactoral, and in our experience over time, have floor and ceiling effects where there are changes, and I think we need to incorporate still the patient's sense that there has been a meaningful change in physical function without specifying the specific instrument.

MS. MALONE: My point is that quality of life is still very, very difficult to measure.

DR. LIANG: It is, but we do it, and it seems to discriminate patients. People have been interviewed with respect to the results of these questionnaires, and they make sense to them, and this is the first time that I think

jam

the patient's view of whether things are better or worse have gotten into the trials like this.

DR. PETRI: Patient global is part of the ACR-20.

DR. [CLINTON] MILLER: The most important thing you just said, in most of the quality of life indicators, in 36 and others, you can distinguish between patients or among patients; that different patients are behaving in a different way.

My experience has been that you don't see changes in the patient over time. Those people that are well adjusted and well supported, et cetera, continue to function at those levels, and they adjust to what their abilities are, and they continue at that same level, but I agree with you, you can tell the difference in patients.

DR. LIANG: Well, they measure change when change occurs. It is all the same.

DR. [CLINTON] MILLER: Well, it is a very temporary thing, very temporary.

DR. LIANG: Well, I think it depends on the condition, but there are longitudinal data that show that their trajectories can go any way that trajectories can go. They can stay the same, improve, and get worse.

I think you can't generalize. You are talking about a lot of disabling.

jam

DR. PETRI: I think that what Dr. Felson pointed out is the changes in the SF-36 are the physical function questions.

DR. FELSON: Well, that is not entirely true either, but that is another day.

DR. TUGWELL: Just to come in and make a plea that any decisions to any questions made be data-driven.

The basis behind the discussions has been we all have very strong views, as we are hearing this morning, but I would hope that any recommendations about what should be asked for in this document should be data-driven. There are a number of ongoing studies right now looking at generic measures and the responsiveness within the rheumatoid arthritis population, and I believe a year from now, we will be able to make a very clear decision, I believe, probably in favor, but right now, it is not data-driven, and therefore, I would suggest this is an issue that could be revisited in a year's time.

DR. PETRI: I take your point, and I don't think we want to be that specific at this point.

I would like to actually address some of the questions. The first question we were asked was how long should a trial be where the claim would be this physical function/disability.

jam

DR. JOHNSON: Should we add another question?

DR. PETRI: We have enough, Kent.

DR. JOHNSON: Should we separate disability from quality of life?

DR. PETRI: I just did that. I did that. Yes. I have rewritten your question. We are just going to vote on physical function/disability. The rest of health status, we are putting aside.

Did you have a question pertaining to this?

DR. LOVELL: I will hold it until quality of life.

DR. PETRI: The motion on the floor was from Dr. Felson. He wanted us to consider two years. Is there a discussion focussed on this point?

DR. LOVELL: Actually, that is what my comment was. I think another way of looking at Kent's statement that quality of life function really has kind of an immediate face validity is that it has very strong market validity, also, a market value.

If I had a drug and I could demonstrate improvement of quality of life, that would immediately open up all realms of possibility as far as marketing.

Using that information, I think we can say this is a very big carrot for companies to go for, and as such, we can use that to get as what Dave suggested, which is to make

jam

the requirement of trials to be longer than signs and symptoms.

That would allow us to get at some of this longer-term information, but also, I think it would allow us to say this is a very much bigger claim than signs and symptoms. It has much more merit and value in the marketplace, and as such, it should have higher standards to satisfy that requirement.

So I think the trial should be longer in this quality of life, perhaps functional assessment, but definitely the quality of life area.

DR. PETRI: Dr. Abramson?

DR. ABRAMSON: I think we are confusing and we need clarity here as to functional assessment/quality of life. I think functional assessment should be viewed separately.

DR. PETRI: Definitely. I have left out quality of life completely.

DR. ABRAMSON: Quality of life as a separate issue.

Functional assessment, I think, depends on the ability of our instruments or tools to discern that in the three- or six-month trial. If there are short-term functional assessment instruments, that should be part of a

jam

three- or six-month trial. That is separate from notions of disability which may imply more structural joint changes, which I think is a third issue, separate from functional assessment and quality of life.

DR. PETRI: Dr. Felson?

DR. FELSON: Let me briefly answer that question. The reason they are included in the core set is they do detect change over brief periods of time, including in nonsteroidal trials over even six weeks. They lag pain improvement by a week or two usually in most of these trials, but HAQ, AIMS, MACTAR, index of well-being, all have been shown to detect these changes.

DR. PETRI: So, David, could you suggest how we could reword this question?

DR. FELSON: I am not sure I have the question.

DR. PETRI: The question, as I phrased it, was we wanted to give the agency some guidance on the duration of a trial to come to the claim of physical function/disability.

DR. FELSON: This is a minimal.

DR. PETRI: Minimal, correct.

DR. FELSON: I think the issue relates to one that Dr. Woodcock raised earlier, which is how do we get at that long-term disabling trajectory of rheumatoid arthritis. It doesn't relate to whether disability is going to improve or

jam

be different in an active treatment group because it is. We know it is.

DR. PETRI: This is getting back to Dr. Strand's point. We are talking about a longevity claim here. So I think what we are asking is what is the duration of trial for a longevity claim.

DR. FELSON: I suggested two years.

DR. PETRI: Let's actually sort of come to closure on this issue. So the question on the floor is the duration of a trial for this claim

DR. LIANG: I think it makes a big difference in terms of minimum. We all like to have more data for longer periods of time, but that is not a standard that you can impose on people. I would like to find out about drugs that improve your function right away.

DR. JOHNSON: In the core set.

DR. PETRI: Remember in the ACR core, we are going to capture that.

DR. LIANG: Why are we discussing this at all? We have function covered. We have already --

DR. JOHNSON: It is inspired by the long term.

DR. WOODCOCK: Could I clarify? The claim of signs and symptoms can be driven by a number of findings in the core set. You could have minimal improvement in

jam

functional ability and get a signs and symptoms claim at three months. You could have no change. You could have a large change, but it can't be totally driven by that because you have to have some of the others.

You could have a very positive change in functional ability and not win in your trials because you didn't -- it is unlikely, but you didn't improve inflammation that much, and so they still had swollen and tender joints.

So I guess part of the question on the table, is it willing to -- is it useful to separate out this domain of disability or whatever it is, or functional ability and separate that and put it as a separate claim?

DR. LOVELL: The part that seems to be sticky here is the physical function. Quality of life is out there, and the instruments are there. People are saying changing that is going to be more difficult than changing physical function.

DR. PETRI: No. I don't think we are saying that. I am just asking that we separate out quality of life completely from this particular question.

DR. LOVELL: We are still stuck with the claim or the indication being functional assessment quality of life.

DR. PETRI: No. I just separated it because there

jam

was no way we could combine the two and have any kind of focused discussion.

DR. STRAND: We are not talking about quality of life. We are only talking about health-related quality of life, and that, I think, can't be separated from disability, nor can it be said to be a measure of disability if there isn't any disability implied, if somebody has an impairment, but they don't feel disabled.

It seems to me what we should be talking about here is something a little longer term than the three months or the six months of signs and symptoms and the more immediate changes in function, but that the two go hand to hand.

Now, it may be hard to measure them.

DR. JOHNSON: Are you saying that you can't measure two-year disability without invoking quality of life instruments? You can't capture it just with the AIMS or the HAQ?

DR. STRAND: Well, in my mind, you can't necessarily imply disability either. We are trying to say maybe stabilization and no progression is as good as improvement, and as Matt said, sometimes that will impact health-related quality of life.

DR. PETRI: Obviously, these two constructs are

jam

related. We are trying to separate the one that we think is most RA-related, which are the physical questions that are on these health status instruments.

David, do you think there is any hope of rephrasing this question or do we need to table it?

DR. FELSON: I think we are at the point where we are not going to get a lot more by discussion. We need to get a sense of the committee. I don't remember how you phrased it. It was nice the way you phrased it.

DR. PETRI: I tried. I wanted to emphasize that this question is one of these longevity claims. You want to look at whether the drug is going to have benefit in terms of the physical function/disability because physical function may to not disable the patient at two years.

Is there any motion to rephrase the question before we vote?

[No response.]

DR. PETRI: So the vote is going to be whether two years is an appropriate trial length for consideration. All those who agree, please raise their hands.

[Show of hands.]

DR. PETRI: May I see a show of hands for those who disagree?

[Show of hands.]

jam

DR. PETRI: I think that carried. Obviously, there are many nuances here.

The next one, I hope we can dispense with more readily. This is where I am going to bring in this quality of life health status. Can a quality of life or health status finding stand alone as a claim? This is where I felt so strongly it could not because a narcotic and antidepressant -- a great number of drugs could conceivably improve the health status of a patient with RA, but have nothing to do with the pathophysiology of a disease.

DR. LIANG: Are you including the physical component of health status or quality of life when you say that?

DR. PETRI: I think I probably am, a patient who is on an antidepressant or CNS stimulant will probably have an improvement in their function.

DR. LIANG: Okay.

DR. PETRI: I wouldn't limit it. If you give someone a narcotic, they may do more.

DR. LIANG: You just separated it out.

DR. PETRI: I promised I would bring it back. So that is what I have done.

DR. JOHNSON: The stand-alone issue actually pertains to the one that you just voted on, also. You would

jam

then have to ask the question to what degree do you need signs and symptoms if you are going to get this two-year physical function disability claim. Do you follow me?

DR. PETRI: I do. Why don't we do health status first and then try to do that.

DR. LOVELL: If a company could come in and show that Prozac for two years truly improves the quality of life for patients of RA, isn't that important information? They are not going to say it changes the signs and symptoms or that it changes the joint erosions, but if a company wanted to come in and really try to test a hypothesis with Prozac, for example, wouldn't that be information that would be relevant to patients and clinicians?

DR. PETRI: Yes. Of course, it is relevant, but it is not going to be the focus of our discussion.

DR. LOVELL: Why not? I mean, we are here, the Arthritis Committee. We are trying to oversee the interests of patients with arthritis.

DR. PETRI: Dr. Chambers?

DR. CHAMBERS: Because I think the question is whether you would label it as an indication as part of a rheumatoid arthritis indication. Prozac was already doing that particular function. It was already labeled and approved because it improved a patient's general sense of

jam

well being, but to subset that in that does the general sense of well being in RA, it is probably not necessary because it already has that indication. So why give it an RA indication?

DR. LOVELL: I see. Thank you.

DR. PETRI: Let me rephrase the question that was on the table before we took the Prozac detour, which was can a health status finding by itself stand alone as a claim.

May I see a show of hands for people who believe that yes, health status by itself could stand as a claim?

[Show of hands.]

DR. PETRI: And a show of hands for no?

[Show of hands.]

DR. PETRI: So there is a consensus there.

Now, we have to go back to the other point, the subpoint, which is can physical function by itself stand alone, so a finding on HAQ or AIMS. Can that stand alone if the study has not met the signs and symptoms claim? Is there any discussion?

DR. SCHWIETERMAN: Can I just ask for a clarification, what you mean by having met the criteria? If we are going to give a certain time point to any of the secondary claims, we might want to consider the simultaneity of other claims or the minimum requirements for those other

jam

claims.

For example, would a three-month signs and symptoms claim be good enough for a 12-year quality of life claim? Perhaps people have comments on that.

DR. PETRI: I think we discussed that. There was a consensus that six months sounded reasonable, and Dr. Chambers brought up the fact that a lot of what we need is not just the efficacy, but the safety data.

DR. JOHNSON: But he means if you are going to have a two-year claim in physical function, is six months enough in signs and symptoms as a co-requirement.

DR. PETRI: I would say yes, but I need the committee to comment.

David?

DR. FELSON: I think a drug should get approved on the basis of its ability to affect signs and symptoms over whatever period of time a company can prove it, and that almost always includes disability.

I think subsidiary to that is if it is approved, then can the company also claim that it affects disability over the long term, and the answer is yes if they show over a long period of time that it affects disability.

I am not sure how that translates into claims, but I guess I would put it subsidiary to an effect on signs and

jam

symptoms.

DR. LIANG: You want to have information on signs and symptoms concurrent.

DR. FELSON: I think that is going to happen. It could conceivably be done by mail, I guess.

DR. LIANG: I actually think people will do whatever it takes to get the data, and I would like to see it concurrently.

DR. PETRI: The discussing is focussing, I think, on the fact that we don't want physical function to stand alone, that it is subsidiary to signs and symptoms, and that the information should be gathered concurrently.

Dr. Abramson?

DR. ABRAMSON: I guess the question here is the metaloproteinase inhibitors. Is this where that discussion should be brought up?

DR. PETRI: We are going to discuss structure next. So I would ask you to hold that for the structure.

DR. ABRAMSON: Well, the question, though, is if those don't have a major impact on signs and symptoms, but prevent structural damage and, therefore, functional ability a year or two later. Is this a problem for this vote?

DR. SIMON: Yes.

DR. PETRI: Dr. Felson?

jam

DR. FELSON: I think in deference to Peter Tugwell's comment, let's comment on the data.

Notwithstanding Lee Simon's earlier suggestion that structural change eventually would realize itself as disability change, there is really no evidence to that effect.

When you look at causes of disability, there are a large number of studies in which long-term RA is followed with disability. Usually, the HAQ is an outcome, but other measures of disability -- no one has really ever been able to show that structural change over time is a powerful -- has a powerful effect on the occurrence of change in disability.

DR. LIANG: I think you are really exaggerating the evidence on the other side, though. Bad function produces bad function, but that is circular. I don't find that satisfying data.

DR. FELSON: People have looked at the effect of structural change on top of that, and no one has ever been able to find a very powerful effect here. There are a lot of things that affect disability. Fatigue does. Muscle weakness does. General well-being does, and probably structural change does.

DR. LIANG: But I think we are talking about the

jam

difference between impairment versus qualitative, semi-qualitative measures, and I think that it would defy common sense that if you suffer whatever it is, cartilage, tendon, muscle, you don't have disability. That is sort of a causal change.

DR. SIMON: And it probably defies common sense, Steve, that if you alter structural change over time that you probably won't alter signs and symptoms.

DR. ABRAMSON: Suppose you had a drug that just was chondro-protective and you had whatever disability and pain that you had, but you didn't progress.

DR. SIMON: But that is a different claim. Somebody that would be chondro-protective couldn't claim that it would alter signs and symptoms.

DR. JOHNSON: Well, sure, it could. It could claim that it altered the natural history of signs and symptoms. It prevents further deterioration in signs and symptoms.

DR. WOODCOCK: I think because we don't have an agent yet, that we don't have this agent, this hypothetical agent, this guideline is not binding, and it isn't intended to encompass situations that we haven't contemplated yet, and I think we can use our common sense and our flexibility.

We are trying to guide people actually right now,

jam

I think, in the main, except for some people who mentioned some things earlier, in agents that are directed primarily at treating signs and symptoms of RA right now, but I think your comments are well taken, and it will be useful to those folk who are developing these more cutting-edge agents.

I don't think they would have to fit into this paradigm we are developing.

DR. PETRI: I am going to rephrase the question, which is that physical functioning is a subsidiary of signs and symptoms claims, and that we would require and strongly recommend that it be obtained concurrently in the same clinical trial.

Could I see a show of hands for yes, there is agreement with this statement?

[Show of hands.]

DR. PETRI: And a show of hands, no, there is disagreement?

[Show of hands.]

DR. PETRI: Thank you.

Now, I have failed in my attempt to keep us on schedule, but we must have a 15-minute break.

[Break.]

DR. PETRI: We are going to take the next two sections that we need to cover before the lunch break. The

jam

first section is going to be on the structure claim, and Jeff Siegel has a presentation.

**Structure (X-ray/Other)**

[Slide.]

DR. SIEGEL: Considerations in the approval of new agents for the treatment of rheumatoid arthritis. We would like to be able to see if there are new agents which, in addition to improving the signs and symptoms of rheumatoid arthritis, also delay or prevent the long-term structural damage and disability that is seen.

[Slide.]

DR. SIEGEL: Delays in progression of radiographic features of disease is hoped to measure something which correlates with long-term disability.

Radiographic findings have been shown to correlate with severity and duration of disease, as well as with functional measures and disability.

However, despite some suggestive studies, there is disagreement in the field regarding whether DMARDs prevent radiographic progression. There is some good studies, but none which are controlled, randomized, large enough studies to be able to reach a definite conclusion.

[Slide.]

DR. SIEGEL: In particular, trials of radiographic

jam

progression have been plagued by methodologic problems, including high dropout rates and the lack of appropriate comparison groups.

[Slide.]

DR. SIEGEL: In a draft guidance document, these are some of the criteria that have been mentioned in regard to prevention of structural damage. Trials intended to support a claim of prevention of structural damage should be a least one year in duration.

A number of outcome measures would be allowable for -- in support of such a claim; in particular, the use of the Larson or modified Sharp or other validated index to show retardation of radiographic progression. Another would be responder or non-responder analysis, looking at the percent of patients with prevention of new erosions.

A third possibility would be to use other measures, for example, magnetic resonance imaging with the criteria of success specified up front in the protocol in order to demonstrate that the criteria for success are clinically meaningful.

In order to obtain this claim, the agent would also be expected to show efficacy in signs and symptoms.

[Slide.]

DR. SIEGEL: Now, one of the quandaries comes up

jam

that there is some evidence to suggest that certain DMARDs may, in fact, prevent radiographic progression, but in long-term trials such as to require to show retardation of radiographic progression, you can't use a placebo control. You have to use an active control.

If methotrexate is, in fact, effective, even though it hasn't been yet proved, a drug which merely shows that it is comparable to methotrexate would not be able to show superiority in a trial.

Therefore, two other trials designs would be accepted as demonstrating efficacy in this regard. One, a drug could show superiority of methotrexate in a head-to-head trial of at least one-year duration, and when I say methotrexate, I mean methotrexate or other standard of care therapy.

The second design would be that the drug, when added to background therapy, showed superiority to background therapy alone.

Now, it is particularly important in these long-term trials to avoid loss of data due to dropouts and protocol violations. In particular, one way to do this is to include provisions for following patients who stopped experimental therapy.

It is critical to keep all of the patients in,

jam

regardless of whether they came off experimental therapy and what other therapy they might be on as a second agent.

A second way of preventing dropouts is to allow some flexibility in treatment options during the study, so, for example, to allow a limited number of joint injections or other measures that are included in the document.

[Slide.]

DR. SIEGEL: We have three questions that we wanted to address to the committee. The first is, are there other trial designs which you might suggest which conform to the logic of the proposed claims structure, besides the two that I mentioned?

Second, if the signs and symptoms claim has been adequately substantiated, would a single trial be sufficient evidence to support a radiographic claim, or do you believe more evidence would be required?

And a third question -- and this is a variant on one which has come up before -- the document proposes the data on clinical responses be collected during the yearlong duration of an X-ray claim trial, but does not specify the clinical efficacy be demonstrated at one year. Do you consider this appropriate?

Let me mention one point here. The claim of efficacy for signs and symptoms could be based on a

jam

placebo-controlled trial, a shorter trial. A separate trial could be done comparing the drug to methotrexate, and the drug may, in fact, be equivalent to methotrexate. So it may not show superiority. So there would be no proof that it is effective for signs and symptoms at one year, but yet, may show superiority in radiographic progression.

So I will stop there.

DR. PETRI: We actually started our discussion of this issue this morning already, and I think I and several other members of the committee felt very strongly that the structure could be a stand-alone claim, having absolutely no connection necessarily with signs and symptoms. It is something Dr. Simon said.

So I think that almost addresses your second and third questions; that we actually think that there are drugs out there in development that may divorce these two claims.

Dr. Schwieterman?

DR. SCHWIETERMAN: Dr. Petri, may I ask you, if new tools -- new tools are being developed for this, and there were to be marginal claims of radiographic progression of some sort, and there was some question, as I think was alluded to earlier, as to how that related ultimately to clinical benefit, what would be your position, then.

I think no one would dispute that. In some

jam

circumstances, you could equally equate structural benefit with clinical benefit, but sometimes we are faced with scenarios where it is not quite so easy.

I can envision, for example, MRI studies in the future or marginal X-ray claims coming in. Perhaps the committee can comment on that. Is this true in all circumstances that you would want this claim to stand alone, irrespective of signs and symptoms?

DR. PETRI: I think the clinician has to make the judgment whether the marginal benefit means that that clinician is going to prescribe the drug to his or her patients, but let me ask the committee for their opinion.

Dr. Lovell?

DR. LOVELL: Well, I think one of the answers to your question is, is it clinically relevant, this kind of X-ray change, does it necessarily, directly relate back to signs and symptoms. It could be just as directly related to functional assessment, for example, or disability.

So I agree completely that signs and symptoms isn't the kind of wherewithal anchoring term here. I think the clinical relevance is important, but we won't know that until we have had much longer experience with these medications, this whole new group of medications that may come up. That will have to be, I think, determined by later

jam

trials, but it doesn't necessarily relate to directly back to signs and symptoms. It may relate to another -- other clinically important parameter.

DR. SCHWIETERMAN: I think that is quite useful, actually. What I am seeking here is, more or less, guidance to situations where we may not have concordance with other clinical outcome measures, but yet, have a radiographic or a structural outcome measure that shows benefit.

My question is, in those cases where there is some doubt, and perhaps there are not cases -- I don't know -- where marginal claims of radiographic or structural benefit are witnessed, ought there be a recommendation to sponsors of these therapies to pursue other trials that look for the things that you are talking about, like functional or other sorts of things, or ought to be a stand-alone claim without regard to those things?

DR. PETRI: Let me ask -- Dr. Chambers has a comment, first.

DR. CHAMBERS: I don't think we were thinking of it being necessarily tied to signs and symptoms. It was tied to something else that showed clinical benefit. The question was can you always tell -- I mean, there are going to be methodologies where there are very tiny changes, and the question is -- we have difficulties sometimes making

jam

that, is it clinically relevant or not.

It was an easy out to say, okay, if you have signs and symptoms, then there was clinical benefit, but if we don't have that easy out or if you don't have a good functional change, would we still take something that was an X-ray millimeter, instead of being wavy, it is now straight, as a claim?

DR. LOVELL: Well, it seems that you are cutting the edge of this technology, and you may inadvertently throw a very promising, potentially important drug out the window by requiring that we -- at the approval level -- I mean, you could require the company to do post-marketing studies to answer that question, but I think you wouldn't want necessarily to put them in the position of having to demonstrate that before they could get the indication for structural damage.

DR. WOODCOCK: Or, you could simply say that the committee could say in those circumstances, they wanted to see more definitive structural change benefits before approval, which is kind of what you are saying, I think.

I mean, if you had two trials that showed on an MRI of what we are raising, a very tiny difference, but it is statistically significant, but it was very tiny and we had no real validation of what that meant to patient, you are

jam

saying you wouldn't want another claim to be shown, but it is potentially -- I think you are saying you might want to say more definitive evidence that it really prevents major structural damage.

DR. PETRI: Dr. Simon?

DR. SIMON: I think this is no different than how we have agonized before in other less-definitive, supposedly, measurement systems. I mean, we are unimpressed by marginal differences in responder indices.

The key issue here is that there is not necessarily going to be this magic thing that is going to do everything. It may well be that we are looking in the future of multiple chemotherapy to take care of patients with this disease, and one thing might be stop erosions. Another drug, unfortunately, might cause new healing that will take place, and a third one might be for signs and symptoms.

I really am fearful that we are going to cripple an evolving technology by being way too demanding in our abilities to ascertain what is going on.

We do have experience in determining what marginal importance means, and I think we have to leave that up to good common sense, but expectations should be written into the protocol or whatever we are talking about to make sure

jam

that people understand we are looking for substantial changes for claim, other than just marginal changes.

DR. PETRI: I think Dr. Simon has summarized the consensus of the committee.

Now, I didn't think that there were any important points about your first question. I thought that the study designs that you proposed were quite logical, but again, let me ask the committee. Were there any comments about the study designs proposed?

Dr. Felson?

DR. FELSON: I don't think it has to be against methotrexate. I hope that was just an example. I would think it could be superiority to any accepted second-line drug.

That would, of necessity, mean that it is superior to placebo, and it wouldn't have to resolve the question of whether the second-line drug comparator is actually superior to placebo. So I think it would be fine.

I would just change the wording of that to say any second-line drug. Now, than, again, we will get into an argument about what a second-line drug is, which I would rather not do, but I think the FDA could probably give guidance there.

DR. WOODCOCK: The only reason for needing a

jam

second-line drug in the comparator arm is an ethical and practical reason.

We saw trials -- this committee saw trials of tenedac where, actually, the comparator was an NSAID, I think, in one of the trials and they did an X-ray trial.

So, in some cases, it may be doable in whatever clinical situation.

DR. LIANG: Question, suggestion possibly. Is there any way that the FDA could establish a Larsen standard for changing -- I mean, a standard test cassette that the companies could use to compare their --

DR. PETRI: This is one of the things that was brought up in the correspondence. Dr. Paulson had suggested very strongly that new technologies are going to be used in this area. He specifically mentioned MR.

DR. LIANG: Irrespective of the technology, I am talking about it is hard to compare study A to B to C because they are using different -- but if you -- and here is something where you could actually imagine a national standard that shows that -- I don't know. You would maybe get a random sample of hand films that have been collected in very structured ways to show what the expected erosions, whatever, and that --

DR. PETRI: Could you get closer to the mike?

jam

DR. LIANG: The companies were basically -- or not companies, but the investigators would use that reference standard to make any assessment of changing structural damage.

DR. WOODCOCK: I think that that type of proposal invitee itself to an academic industry and FDA collaborative effort to do something like that.

DR. PETRI: Yes. Dr. Johnson?

DR. JOHNSON: I would like to plead for a little more conversation. I think at least it sounds attractive on the surface, but underneath, it gets us nowhere.

If we are not going require concomitant claim in circumstances where there is a substantial difference, then we have to figure out what substantial means, and nobody knows, and nobody knows how to know because we don't have the data.

We have got an invalidated surrogate here. It is different from blood pressure. We had some blood pressure trials where mortality were endpoints and a hell of a lot of epidemiology, and now we have got a little epidemiology and an unsubstantiated endpoint.

You could say you study at Phase 4, but Phase 4 studies have a problem with being as rigorous as Phase 3 studies, and we may still not get the answer.

jam

DR. PETRI: Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I think what Lee said, I agree with him.

DR. PETRI: We need some help getting the microphones turned on at the committee table.

DR. FERNANDEZ-MADRID: I think I agree with him. In answering that question, I think a very significant difference in signs and symptoms concomitant to a marginal difference in the structural parameters will not validate those. I don't think this would make it.

DR. PETRI: Janet is whispering to me that she thinks that we have come to a reasonable consensus.

DR. SIMON: I want to ask one question.

DR. PETRI: Yes, Dr. Simon.

DR. SIMON: In what has been written or the question that has been asked, it is about second-line therapy, and I am a little concerned in David's illusion that, in fact, that should stand. Your comparator would be a, quote/unquote, "second-line therapy," and if that is written into the document, that is very constraining.

DR. PETRI: It was suggested as one possible study design. This is not supposed to be all-inclusive.

DR. SIMON: That is fine. I just want to make sure that terminology would not be restrictive.

jam

DR. FERNANDEZ-MADRID: It is not written in there, and there was no intention to write it.

DR. PETRI: We need to move on. Our next section is going to be on major clinical response. There are going to be several participants, including Kent Johnson, David Felson, and William Schwieterman.

There was a question of what was the consensus. We felt that the structural claim could stand alone.

We are not voting on everything because of the time constraints. If I feel there is a clear consensus, the Chair is announcing that.

We felt there were several study designs that would be appropriate to determine that claim, including trials against another second-line agent, and we wanted to leave it open to industry, academia, and the FDA to collaborate on determining what the outcome measures would be for a structural claim, such as some consensus on a Larsen score or MRs.

### **Major Clinical Response**

DR. JOHNSON: I am going to be brief. We have touched on major clinical response.

[Slide.]

DR. JOHNSON: This was inspired by a lot of people who wanted a shot in the arm for drug development for

jam

rheumatoids you can't remit because they have got deform disease and they are always going to have some pain due to the deformity.

[Slide.]

DR. JOHNSON: In addition, we wanted it to be a major, major hurdle, to try to get away from sort of incrementalism, something not quite akin to remission, but something pretty close to it.

[Slide.]

DR. JOHNSON: That is about as far as we got because we can't figure out how to define it. Obviously, we really need to recognize in some way what exactly is it about these patients that does prevent them from remitting at least intuitively.

[Slide.]

DR. JOHNSON: You can think about various ways of going around this. You could just go your seat of your pants and try to get sort of a consensus of rheumatologist or patients or both.

Included in this, you might actually incorporate some novel things such as No. 4, connubial by MRI or even connubial biopsy if you were really determined to find out. I am sure there are sampling areas in all of that with those approaches.

jam

The other approach is what we have asked Dave Felson and Jennifer Anderson to address, and that is do they have a data bank that could help us define. If we told them what proportion of the historic DMARD patients, you know, what top fraction we wanted to have derive our definition of this major response, how would they do that, whether it is a 1-percent cutoff or a 10-percent cutoff or whatever.

So I am going to turn the floor over to Dave, and then we will go on to questions after that.

#### **Toward a Data-Driven Definition**

DR. FELSON: Kent called me and he said, "Well, Dave, how might we go about defining major clinical improvements?" We talked about step thresholds earlier, presumably the threshold above the ACR improvement level, and Kent suggested to me that the way they had been thinking about it at the FDA was the top 10 percent or 5 percent or 1 percent of improving patients on second-line drugs.

Let me actually redefine that more precisely: the top 1, 5, or 10 percent of patients who have received second-line drugs in trials in terms of their level of improvement.

So that was the sort of approach that we adopted. In addition, I should note to you that as you have probably seen at the ACR meetings and perhaps even now in the

jam

literature, the ACR 20 percent has been -- a new threshold has been added in many papers and many presentations, and the ACR 50 percent which we never developed -- and we are sort of wondering whether that is an appropriate threshold.

My approach in the next 5 to 10 minutes, I have actually culled my transparencies a couple of times because I knew time was really limited here. I am going to start with the ACR core set, to review the definition of improvement development, very briefly, and the reason I am doing that is not because you need to hear about it, because we have already been talking about the ACR response criteria, but to remind everyone, the criteria validity that we used in developing the improvement criteria, so that we might have a similar approach to the development of a definition of major clinical improvement, and then I am going to try to do this sort of in a data-driven way.

[Slide.]

DR. FELSON: This will go briefly, I promise you, and it is not meant so that you can read all the detail. It is meant to impress you with how much detail there is.

You can see here on the left there is the ACR, ILAR, ULAR core set, which has the seven parameters we talked about, chosen because they were nonredundant, all sensitive to change, and sampled broadly from the content of

jam

improvement that occurs in rheumatoid arthritis patients when they improve.

Each of them is operationally defined on the right, and you will notice, by the way, that when you look down at No. 6, there are a variety of different definitions of patient assessment of physical function that at the time corresponded to those instruments which had been proven to be sensitive to change in RA trials. I think that kind of selection is probably appropriate.

[Slide.]

DR. FELSON: Just to remind you, there were a whole set of validity concerns that we addressed in developing the core set and also in developing the definition of improvement that I would recommend some of these for thinking about major clinical improvement, and I will try to briefly suggest how we might do that.

They include face validity. Is the definition of major clinical improvement credible to the rheumatologist in the audience here? Does it identify patients who we would all agree have had major clinical improvement? Is it comprehensive and discriminate validity? Is it sensitive to change? Do outcomes define as major clinical improvement?

[Slide.]

DR. FELSON: This is the process that we used in

jam

coming up with the ACR definition of improvement, and I put it up more to remind everyone here that in coming up with a definition of major clinical improvement that some of these elements ought to be appropriately discussed or pondered.

We surveyed rheumatologist using paper patients and asked them based on a series of patients listed in trials who had improved by various degrees who of these patients do you think improved and who didn't, and based on the rheumatologist' perceptions, we selected a number of definitions of improvement that corresponded well to clinical impressions of improvement.

I might suggest that a definition of major clinical improvement ought to incorporate the same kind of question to rheumatologist.

We analyzed survey data.

[Slide.]

DR. FELSON: We had 17 definitions left, and then we used clinical trials data to try to figure out which definition of improvement we had selected from rheumatologist' impressions. That has separated effective second-line drugs from placebo.

So there was both a consensus and an impressionistic process here, a survey process, and there is a data-driven process that related to the analysis of

jam

clinical trials.

[Slide.]

DR. FELSON: We selected the definition which worked best on all of these parameters, and it turns out to be the definition of Wc. What you see here is the discriminate validity analysis of the ACR improvement effort, and each of those letters stands for a different definition of improvement that we tried.

You notice there are a lot of different D definitions. Those are different families of improvement definitions. Different W definitions are other different families. Wc was the one we wound up selecting for a variety of reasons, in part, because it had better discriminate validity.

Those bars you see arching down from left upper to right lower are isopower curves. They tell us how powerful each of these definitions is in discriminating between treatment and placebo groups. So you will see Wc tends to be fairly powerful. Some others, for example, I-7, way at the top left, is perhaps the most powerful. It is a straight index of all of the seven ACR core set measures.

[Slide.]

DR. FELSON: That was a development sample. We then tested in another comparative trial. So this is the

jam

definition of improvement.

Now, I am going to leave this up because I want to show you how we then tried to address Kent's question to us. The ACR preliminary definition of improvement for RA trials, one we have talked a lot about here, requires that a patient entering a trial experience both 20 percent improvement in tender joint count and 20 percent improvement in swollen joint count, and they also must experience 20 percent improvement in three of the five following core set measures listed there.

So the approach we took to defining major clinical improvement that was suggested by Kent was to raise the bar higher using the same rule.

So, for example, when I talk about a 50 percent threshold, what I am going to be characterizing is a patient will meet a 50 percent threshold if they have experienced a 50 percent improvement in both their tender and swollen joint count and a 50 percent improvement in three of the following five.

[Slide.]

DR. FELSON: Now, the trials that we have got data on are listed here. Before I go further into the trial data, I want to make a couple of very preliminary comments.

First is to offer sincere thanks for the help of

jam

three statisticians who in a period of a few days helped me put this together: Jennifer Anderson, who no longer works with us; Mike Levali, who now does; and especially George Wells from Ottawa, who I really relied on for the cyclosporin methotrexate combination trial data.

I also want to mention that while the original draft from the FDA suggested that an 80 percent improvement threshold might be an appropriate initial guess -- and Kent, actually, in a phone call suggested even 90 percent -- as appropriate initial guesses, we found no patients who improved that much. It is just like no patients who were in remission in these trials. So it was not terribly informative to use those thresholds.

So what you are going to see is thresholds extending from 20 percent up to 70 percent, where there is information.

[Slide.]

DR. FELSON: So I would suggest, as we begin to look at these things, that we consider three issues. One is the percent of patients who have been proved on known efficacious treatment. That is what Kent asked us. Another is the discriminate validity of any threshold.

In other words, if we were to use that definition of "major clinical improvement" in our trial, would we have

jam

any power. If we were to add it to the 20 percent improvement criteria, would we have any power, and the next is face validity, and I will present some data on face validity, which is among the rheumatologist here and people who are used to seeing patients here, does this definition correspond to what you and I would think if we saw a patient and said, look, this patient has experienced major clinical improvement, do these definitions correspond do that, and I will show you some data about what kinds of patients each of these thresholds identifies.

So the trials that we are going to be talking about are pooled cooperating clinics trials of second-line drugs in which I will pool together three trials, the methotrexate trial versus placebo, a D-Penn trial in which 500 milligrams was used versus a low-dose D-Penn, 125 milligrams versus placebo, and then a gold trial, a gold auranofin versus placebo, and I will be looking mostly at the gold arm, a methotrexate versus auranofin trial done by Mike Weinblatt a multi-center trial, similar doses of methotrexate, a dose of auranofin, we will talk about, and then a trial you are familiar with from yesterday, a combination trial that was in the New England Journal.

Now, I think the most salient difference between these trials is this one, RA patients, and you will notice

jam

that No. 1 and No. 3 look at RA patients, mean duration of disease, 10 years, and analyses we have done and others have done, as I have suggested earlier, pointed repeatedly to the idea that early-disease patients are more likely to respond.

So what you are going to see is higher response rates in this trial, even using essentially the same treatment regimens. So duration of disease is the predictor of response here, and then you will notice there is some heterogeneity between trials that is explainable in part by that.

[Slide.]

DR. FELSON: So here is the CSSRD data, and I am not sure I can see it from here. What you can see is this is the ACR 20 percent. This is the ACR definition of improvements.

You will notice when the strong drugs pulled together, the rate of improvement was 40 percent. The placebo rate of improvement was 8.5 percent. The chi square distinguishing between the strong drug and placebo is 36.9, highly significant, one of the reasons why the ACR threshold definition works. It gives you a lot of power in discriminating strong drugs from placebo.

Now we will move up the ladder. You will notice because of the small numbers of strong treatment that

jam

reached this threshold, you lose a lot of power. So you don't get the discriminating ability, but the good news is that none of the placebo-treated patients reached that threshold, okay, which is sort of reassuring.

[Slide.]

DR. FELSON: So this is now the methotrexate trial, and you will notice immediately that the rates of improvement are greater using any given threshold, and remember, this is an early-disease trial. So here are methotrexate-treated patients, and you will notice now that 65 percent of methotrexate-treated patients reached improvement based on the current ACR definition. If you use a 50 percent threshold at 25.3 -- I know it is hard to see -- a 60 percent threshold to 18.5 percent and an ACR 70 percent threshold, it gives you 9.2 percent of methotrexate-treated patients, and believe it or not, 6 percent of auranofin-treated patients, which was surprising to me.

You will notice once again that the higher up you move on these thresholds, the more power you lose. So that would suggest that it doesn't matter what threshold we choose. It would not be advisable to suggest this threshold at the current time as the single determiner of whether a drug is efficacious or not. It is just not powerful enough.

jam

[Slide.]

DR. FELSON: Once again, my sincere thanks to George Wells. There are two groups in the cyclosporin methotrexate. So, remember, this is a combination therapy trial in which methotrexate-stuck patients are randomized to cyclosporin or placebo, and these are the ACR improvement rates in these two groups.

I would focus here. This is the strong group on the left side, those that received two drugs, and the ACR improvement rate is 45 percent, using a 20 percent of the ACR criteria. If you look up to 50 percent, it is 22.5 percent, 60 percent, 5.6 percent, 70 percent, 1.4 percent. So you can see, once again, the numbers drop dramatically as you increase the threshold.

As I pointed out earlier with Kent, the 80 percent threshold is simply uninformative.

[Slide.]

DR. FELSON: That summarizes the data from the three trials at the thresholds that I think are the most reasonable.

These are the CSRD trials on the left column, methotrexate only from the methotrexate auranofin trial in the middle, and the cyclosporin combination.

Now, the typed data are data from both arms. The

jam

written data that I got from George Wells yesterday, I wrote in, are the combination therapy patients only, percentage improvements. So I would compare 9 percent, 35.3 percent, and 22.5 percent. I would compare the 60 percent threshold. It would give you 2.6 percent, 18.5 percent, and 5.6 percent, and the ACR 70 percent would give you 0.6 percent, 9.2 percent, and 1.4 percent.

So those are roughly the rates of improvement, major clinical improvement that one would see, depending on how one defined major clinical improvement.

[Slide.]

DR. FELSON: The next question is do any of these definitions correspond to what you and I would call major clinical improvement in a clinical sense.

So let me give you some examples that I thought were characteristic examples of patients reaching each of these thresholds, before treatment, after treatment. This is examples of patients who experienced 50 percent improvement, but not 60 percent improvement, and half of the patients who experience 50 percent improvement do not experience 60 percent improvement. So there are a lot of patients here.

Tender joint count, 36 before treatment, 14 after treatment, 38 swollen joint count before treatment, 19 after

jam

treatment, pain drops by more than half, patient global drops by half, MD global does not drop by half, grip strength does not increase doubly, and sed rate drops by well over half.

Actually, there is a caveat that I have to tell you about because I have forgotten to mention it, and it is an important caveat. The CSSRD data that you are going to be seeing here all uses grip strength as a substitute for HAQ or disability because we didn't have HAQ or disability measured in these trials. That is an important limitation.

Our data now coming from trials, including the methotrexate cyclosporin trials suggest is it not a limitation that has affected the validity of any of our results.

We knew when we did this that grip strength and physical disability were closely correlated, and grip strength was not a bad substitute for HAQ, and that is why we did it.

Another patient, 50 percent improved. You will notice these improved considerably, pain a lot, MD global just by half, grip strength not much at all, and sed rate barely by half.

[Slide.]

DR. FELSON: Let me give you an example of 60

jam

percent improvement who don't make 70 percent improvement. Now you will begin to see, I guess in my clinical view, these patients -- I could not find a patient in this group who I was thinking didn't really improve in a real major way.

25 before, 6 after, 20 before, 4 after, 13 before, 9 after in terms of pain, and then major improvement in patient global, major improvement in grip strength, major improvement in sed rate, and you can see sort of a similar phenomenon going on with patient two.

[Slide.]

DR. FELSON: Then, let me show you the 70 percent improvers. There are a few of them, but their improvement is impressive.

Here is a tender joint count going from 20 to 2, swollen joint count going from 11 to 1, pain going to zero, global going to zero, MD global going to zero, grip strength dramatically increasing. Every one of these patients was like this. This is not selected.

If you choose a threshold of 70 percent or greater, they all have very impressive improvement. So I think that is an important message.

[Slide.]

DR. FELSON: I think I am done. I am sorry to

jam

take so much time.

Let me just give you a few of my observations. Using ACR thresholds, it has face validity. In other words, it corresponds to clinical impressions of major clinical improvement. I think we can, in other words, use Kent's suggested approach, which is to come up with a new 50 percent, 60 percent, 70 percent, and it works fine.

Using ACR thresholds of 50 to 70 percent would not have sufficient discriminate validity for uses of primary outcome measure. There is not enough power there. So it would be a supplementary threshold in addition to the ACR definition of improvement.

Defining major clinical improvement as greater than 50 percent improvement may identify some patients whose response has not been extremely impressive. That was my impression looking at these data. Yet, it is easiest to use of the alternatives. It is really easy to use.

A lot of people already are using it. You can figure it out in your head. It is very easy to use. It gives you more power than thresholds that are higher.

The downside is that some of these patients may not have numbers that you would say, wow, is that patient improved, but the plus side is that it is easy to think about. It is easy to use. People are already using it, and

jam

it actually has got more power than any of these other measures. It is more informative.

I think further work is needed in face validity, and probably further work is needed on an issue we talked about earlier today, which is how might one analyze data in which there are several thresholds for response at different levels because one then could incorporate that into a single outcome measure, like an ordinal outcome measure, 012 response or 0123 response, and get even more power than what we have got now.

Thanks for the time, Michelle. I am sorry to take up so much time.

DR. PETRI: What I am going to suggest is that Dr. Schwieterman give his presentation, but after his presentation, we will break for lunch. The discussion at this point will be after lunch.

### **Complete Clinical Response and Remission**

[Slide.]

DR. SCHWIETERMAN: My presentation is going to be very short. Actually, we have had quite a bit of discussion about this topic already. So it doesn't need a whole lot of introduction.

I have five overheads in total, including this one.

jam

Let me just say from the beginning, though, that I think some issues have already come up in this area regarding the utility of these two claims. Some have expressed some doubt whether it is useful to have these two claims, and if I heard it right, it was because perhaps of poly-pharmacy, difficulty distinguishing the two, clinical utility and so forth.

So I would like to continue that discussion that we tabled early in the morning.

[Slide.]

DR. SCHWIETERMAN: As we defined it in the document, we had two claims, one entitled complete clinical response and the second entitled remission. Both are the same thing except that one requires ongoing drug therapy, that is complete clinical response, and remission was the same thing except off drug therapy, and both of them were defined by remission by ACR criteria and radiographic arrest as demonstrated over a continuous six-month period.

It was a feeling of the committee that there would be value to having a hierarchal system like this because patients -- because a claim of having a response while off drug therapy would perhaps be viewed as better than requiring continual therapy, but this is open for discussion.

jam

[Slide.]

DR. SCHWIETERMAN: I am just going to remind you of the 1981 ACR criteria that defines remission. All of you are familiar with this. It is less than 15 minutes of morning stiffness, no fatigue, no joint pain by history, no joint tenderness or pain of motion, no swelling of joints or tendon sheaths, sed rates below 20 in males and below 30 in females.

[Slide.]

DR. SCHWIETERMAN: There were a number of useful comments earlier in the morning that I think we can continue to explore about trial design considerations for these particular endpoints. I will just list several of them here. There are categorical endpoints that need to be prospectively defined if they are going to be used.

We recommended that trials be at least one year of duration in the document because time shorter than that invited problems with characterizing the durability or the legitimacy of these claims, but we can have further discussions about that as well.

I think several people remarked that it is important that the baseline disease status, including signs and symptoms and structural damage, be adequately characterized, and I would just add to that. I think if we

jam

are going to include these claims in the guidance document, that perhaps we might want to add a wording, as some have suggested, about what are meaningful inclusion criteria that you could use to get these particular claims, given that different patient populations might have different propensities toward remission or complete clinical response, and therefore, there would be different meanings to the types of trials being conducted.

I mention in the last part here that, obviously, there would need to be adequate evaluation during the trials to guarantee that, in fact, the definitions of complete clinical response remission had been met through blinded assessments, regular intervals, and so forth.

So I have a simple question at the end, which is good, because there has been a lot of discussion already.

[Slide.]

DR. SCHWIETERMAN: I would just like for the committee to comment on the usefulness of these two claims, complete clinical response and remission, and secondly, to comment on how trials might best be designed to study these endpoints.

DR. PETRI: Thank you.

So we are going to reconvene at 1:15 to discuss this point.

jam

[Whereupon, at 12:11 p.m., a luncheon recess was taken.]

jam

A F T E R N O O N S E S S I O N

[1:15 p.m.]

DR. PETRI: We are going to finish the questions that we did not have a chance to start before lunch, and I want to refer everyone to the second page of questions.

We have been asked to address at least three questions about the major clinical response claim and the complete and remission claims.

The first question is: What is the most appropriate way to define the major clinical response? Obviously, this is something that Dr. Felson began to address.

Dave, if I could ask you, do you want to phrase the question for both discussion and vote on this?

DR. FELSON: Let me start out by saying I am nervous specifying a threshold here at this meeting, and I am wondering if we have the opportunity to sort of ponder this a little bit more in another forum.

I would say that I would suggest there is a need to define major clinical improvement separate from the ACR improvement criteria, and that one can define it using the ACR rules, and that that definition ought to be based on clinical judgment as to what a major clinical response is and the likelihood of response, and based on given

jam

therapies.

I am not sure what the right threshold choice is right now. I think it is somewhere between 50 and 70 percent, and I can't -- I think we need a little bit more work, but if the FDA is really intent upon sort of finalizing this, I think we can guess, but I would rather have some rheumatology input from the committees and stuff. I am not sure.

DR. PETRI: But isn't it a reasonable first step to suggest that probably it should be somewhere between 50 to 70 percent?

DR. FELSON: Yes. I think that would work fine.

DR. PETRI: Because I think that kind of information would be helpful, even though we all realize that wouldn't be final.

Let me open this up for discussion for other suggestions.

DR. JOHNSON: You would have to heuristically think to yourself do you want this -- and I reveal my bias here -- do you want this definition to capture 10 percent of the patients that were in Weinblatt's study.

Even if you used the 70-percent cutoff, you would capture 5 to 10 percent of those patients of the methotrexate versus -- well, 5 auranofin and 10 methotrexate

jam

or something like that.

Part of all this is what we are intuitively trying to aim for as clinicians.

DR. PETRI: I think what we all recognize, Kent, is that this is going to change as we get better drugs, and 50 to 70 is probably not going to be sufficient. We will be asking for 90 as a major clinical response. So I think this is going to be in evolution as our drugs improve.

Dr. Lovell?

DR. LOVELL: Looking at the few patient profiles that were shown, it looked like at 50 percent, there were patients who had still enough disease that I would guess their X-rays would continue to worsen their functional assessment, but when you went to 60 to 70, those patients were dramatically better, and you just got the feeling that those patients probably would maintain their level of benefit for a long period of time if you could keep them there.

I mean, there was a dramatic difference between the patient profiles of 50 and 60 percent. So, if you are looking for this major clinical benefit to be one that would be translated to long-term clinical improvement or stability, then the 60 or 70 percent cutoff would be more like it, I think, to reflect that.

jam

DR. PETRI: So, unless there is further comment, let me phrase this as a question for a vote; that we think it would be reasonable to have the ACR 70 percent as a current definition for a major clinical response, understanding that this will evolve over time.

Those who are in agreement, would you please raise your hands?

DR. FELSON: Before you vote, can I --

DR. PETRI: Just in the nick of time, Dave.

DR. FELSON: Yes. I am nervous about 70 percent. In all of the second-line drug trials, including cyclosporin and methotrexate combination trial which I think we all recognized showed a fairly impressive response, the number of 70 percent responders is trivial or none. So I think we have got to be careful.

I realize that is going to change, maybe.

DR. LIANG: But that is the truth.

DR. FELSON: That may be clinically correct, Matt, but I think we also want a definition that is informative, okay? Putting a bar at a level that doesn't get any patients above that bar is not statistically informative.

DR. PETRI: Combination cyclosporin and methotrexate, you are going to get some people up there, right?

jam

DR. JOHNSON: No. You only got one or two, but in a three-year duration rheumatoids, in Weinblatt's study, you have got five in auranofin and 10 in methotrexate or something like that. So it is going to reflect back to your demographics or your patients.

DR. PETRI: Dr. Simon, first.

DR. SIMON: I mean, one of the big arguments here is what is driving this train. Is it what we would like to see as the major clinical response, or is it what we have accrued? In fact, it is what we like to see. It doesn't necessarily mean it is achievable today with today's technology.

DR. FELSON: Let me just make a comment. You asked the right question. Let me say that there is another thing driving this decision. There are two issues. One is what we would like to see and what we would call major clinical improvement clinically, and the other is what is efficient, what is going to help us figure out drug efficacy reasonably, what is going to make it reasonable for a company to come in with acclaim of major clinical improvements.

A 70 percent threshold at this point would not be a reasonable bar because nobody would be able to demonstrate clinical improvement probably even of the anti-TNFs and

jam

others that might be very impressive.

A 50 percent bar has enough patients behind it, and it shows enough clinical improvement that it would be feasible and efficient to use as a bar.

So I think there are two related issues that there is a tension between, and I think I am nervous because of the tension. One is what we call clinical improvement, and the other is statistical efficiency issues.

DR. SIMON: I would be incredibly uncomfortable personally to use the descriptor, "major clinical response," unless there was a major clinical response clinically, how I would actually interpret that.

It is measurable. Unfortunately, it has not been demonstrable in this particular construct, but I certainly see what you are saying, David. I would still argue for the higher rather than the lower bar.

DR. WOODCOCK: I would like to ask one question of David. That was a single time point achievement. The way it is written in the guidance right now, it requires a continuous six-month achievement of this state, number one, and number two, of course, it is going to require not only that patients achieve it, but there be a statistically significant greater proportion of patients achieving that state compared to the other comparator arm. So that the bar

jam

is fairly significant, as it is written. Now, any of that can be modified, but it does have a duration component in it right now.

DR. PETRI: Dr. Luthra?

DR. LUTHRA: I was just going to make a comment regarding the same issue. I think you have objectively identified a group of patients who truly have a significant response, and if we start to compromise on that, then I think we should really not even mention this thing.

If we are going to say that there is a group of patients who have a significant clinical response, then that is the criteria that we can judge.

I agree with Janet's comment that those patients should remain in that state at least six months. I think we will have to have very rigid criteria. If we start to compromise -- because what you are doing, David, now is thinking about the current state of affairs and trying to find patients that you will fit into that category and asking industry that whatever current drugs at there should be able to find some patient in that. Maybe that is not achievable, and if you are going to do that, then I think we should not even really mention this whole point of complete clinical response.

DR. PETRI: There is a comment at the microphone.

jam

DR. LACHENBRUCH: Lachenbruch, CBER.

I think what I am hearing Dave say is that we are going to be looking at products that are going to have a very low power in achieving the major clinical response.

My suspicion is that if a company would be rather hesitant to gamble on a 10- or 20-percent chance of getting that outcome, if a similar -- if the product had, say, a 90- or 95-percent chance of achieving a signs and symptoms. So we are perhaps left with saying they win the signs and symptoms, but as a matter of fact, they won so well that they ended up -- we will give them the major response, and that creates some statistical problems that I would have a concern with.

DR. PETRI: Dr. Abramson.

DR. ABRAMSON: Yes. I think we may have trouble defining major with our different opinions in these response, but to the extent that it is a valid measure, we are giving some credence to the ACR-20 for efficacy, which I think in the office is a much less impressive response than on paper.

I would encourage us to look at whether we should look at the ACR numbers as useful markers and use them, ACR 50, 20, 70, and not get into the definitions of whether that is major, significant, or complete.

jam

I think that will help us also compare drugs. When we talked about comparisons, if we really could look at this drug that causes a response of 50 percent in this number of patients, I think that is useful information to the clinician.

So there will be questions about the validity of the ACR criteria of improvement as a bench mark, but we are de facto using them this way, and I would like to think about knowing or making public in these studies the ACR 20, 50, 70, or some bench marks. I think that is useful information.

I was very impressed with David's presentation and how the percentages of patients in each of these groups sort of flowed down each of the brackets in a very impressive way.

DR. PETRI: Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I think the data David presented was very impressive and very convincing in terms of the relationship between these ACR criteria and what happens in real life.

Up to ACR 50, really, we will not be convinced that these patients had a significant clinically important improvement. That is, they still have maybe 50 percent of improvement in pain and swelling, but they were very

jam

symptomatic, and this is not really a major response.

So I would think that if we are going to label this a measure response, we should stick to the 50, 70 level.

DR. PETRI: In order to bring this to a vote, I am going to have to, I think, have two votes. So the first vote is going to be used at the 70.

DR. LOVELL: Well, could I ask a question of David? At this point, you have given us kind of a very quick, but very impressive data analysis. How comfortable do you feel with shortcutting the next steps in the validation process that went into the making of the ACR 20? We are still very early in that step validation process as you showed you went through for the ACR 20. How comfortable do you as an investigator and other people who are familiar with this field feel about shortcutting the next iterative steps in the validation process?

DR. FELSON: Well, if you haven't gotten the idea already, Dan, I am pretty uncomfortable. I think it is really hard to sit around the committee here without a lot of thought and cogitation about what these patients look like and input from other rheumatologist who see a lot of RA patients and other data from trials where responses aren't as impressive as you might hope for ideally in practice to

jam

come up with the right threshold. I don't think that is an off-the-top-of-your-head look at a couple of patients on transparencies.

I think that is not a way to come to a thoughtful appropriate decision. I think it would be better to have some discussion of this and more time spent.

DR. LOVELL: And more data.

DR. FELSON: Maybe more data, too.

DR. LOVELL: Maybe we could vote and say that we think this kind of escalation of the ACR criteria at higher percentages is a valid way to go about it, but that further study needs to be done.

DR. PETRI: I think that is a given. Remember, we are not saying someone has to do it this way. We are suggesting an approach.

Let me rephrase the question. What we would like to consider a definition of in major clinical response is capturing that 5 to 10 percent of the best patients in the historic DMAR trials and that one approach that might accomplish this is to use the ACR 70 as the cutoff, one approach. I am sure there are going to be many others. So if we could actually come to a vote on this one.

Those who agree with this statement, would you please raise your hand?

jam

[Show of hands.]

DR. PETRI: Those opposed?

[Show of one hand.]

DR. PETRI: There is one opposed.

The audience is not allowed to vote.

DR. LOVELL: The part that makes me uncomfortable is that we jumped on the 70 percent cutoff, just bam.

DR. PETRI: As I promised, if that first vote had been a nay, I would have gone on to 50 to 70, but since that first vote passed, I don't think it is necessary to look at other cutoffs.

This is one approach. We are not binding anyone. We are suggesting this is a reasonable approach.

DR. TILLEY: I have one last comment.

DR. PETRI: Yes, Dr. Tilley.

DR. TILLEY: It seems to me that, again, we are talking about two different things. I think our clinical people are talking about individual patients and what they would call a major response, and then what David has been talking about is a trial design criteria. So I don't think what we are voting on is what we would call an individual patient. We are voting on a criteria for a different kind of trial, and so then the question is, do you want to do a different kind of trial with a higher bar, and if you do,

jam

David's data suggests that 70 is too high. If you want to define the kinds of patients who you consider clinically to have major response, then the bar is probably quite appropriate.

DR. LOVELL: The ACR 20 was validated both statistically and by the review of individual patient data, and I think one could do the same thing with the ACR 50 or 60 or 70 as part of a validation process. So that, when we get done at the end of the day, we would have the best of both worlds, like we do at the ACR 20, which would be patient impressions that people could be comfortable with as well as some sort of statistical validation, or at least we could address those two questions and see if we could answer them both because I think they are both very important, and we just can't kind of say right now that the ACR 50 or 70 is going to address both of those questions because I think they are both very important.

DR. PETRI: A comment from the audience.

DR. STEIN: Stein, FDA.

Perhaps I misunderstand the situation, but as I see it, the percent response is with respect to baseline, and baseline is not a blinded point in time. Therefore, you can improve the response by simply inflating the baseline, and so I see that as an impediment to judging what the true

jam

state of nature is.

DR. PETRI: David, did you want to respond to that criticism?

DR. FELSON: Actually, I don't. I wanted to address a different issue, which Janet Woodcock brought up, which I think is important to think about, and that is time. We have dealt with time a lot today, both not having enough and wondering about what patients need certain things.

I would say that if you -- thresholds if improvement that we have developed and that have been discussed here today have been based on end of trial, single point in time measures. If one begins to demand that major clinical improvement or any of these other measures be persist over a certain period of time, those percentages start dropping way off.

So any threshold -- 70 percent threshold, if it were barely reachable, it now becomes unreachable. A 50 percent threshold modestly reachable becomes much more difficult, and I think part of any process of committee work or cogitation or thought about this has to also bring that issue in as to whether a clinical response, defined clinically, ought to be based on the time of that response and what that does to the efficiency and design of a trial. I think those are also relevant issues.

jam

DR. WOODCOCK: I think we can take all of these comments now under advisement. It has been very helpful to us.

We were trying to develop a claim comparable to remission, but for patients with fixed structural problems. It also appears there has been a lot of discussion around the table about developing a claim that really reflects a major clinical response.

We will take all of this under advisement, all of this advice. I think we have had a lot of device on this particular and very difficult issue.

DR. PETRI: I think there was one other question that we should actually discuss briefly, and that is on the second page. It was about whether the complete clinical response remission claims should be kept separate, the idea that remission meant off drugs. We discussed a little bit this morning what does off drugs mean. Does that mean off the drug that is being tested in the clinical trial? Does that mean off all drugs such as off NSAIDs?

Is there any discussion about this? I think, Matt, this morning, you had a strong opinion about it.

DR. LIANG: I thought that Steve was suggesting something which we should revisit, which is to eschew these things, these adjectives and just report things as ACR

jam

percent, and then I think it is important that people describe in their data whether the person is still on the drug or not. I mean, I think that is informative information that clinicians and patients want to get, but I would get rid of these sort of adjectives, these loaded words, and just report percent improvement in whatever way we eventually do it. I thought that was Steve's --

DR. PETRI: That is really getting back to our discussion this morning that there are these tiers within signs and symptoms.

DR. LIANG: I am saying to stop asking this question. I don't want to hear it again. I think the carrot still there if you say whatever, ACR 70 or ACR 73. The carrot will always be there. I'd like to hear that rather than --

DR. PETRI: Well, then let's bring that to a vote. The motion is we should try not to use these specific claims, and instead, ask industry to report responders in a continuous way.

David?

DR. FELSON: Now we have to actually move from the statistical, which that is good for, to the clinical, which that is not so good for.

I think there are clinically definable states of

jam

improvement and major clinical improvement that I think we can all come to some consensus on that do reflect a claim that a company might have for a particular compound, and I think those are valuable, and I think those can be defined as different thresholds of ACR improvement. I think we should use them.

While I think one -- I think the other thing is we are going to start throwing them around. In fact, we are already throwing them around. We are already saying certain drug produces X percent of patients get ACR improvement on a certain drug, while another drug, using the same kind of patients, gets X percent improvement, X percent ACR improvement, and I think that is a very valuable constant.

People do it in oncology all the time. X regimen produces X percent partial remissions. And that is what we want here. We want to move toward that. We want to move toward major clinical improvement definitions where we say the X treatment causes X percent of patients to go into major clinical improvement. That is discouraged when one goes to Steve's idea, which is how many -- you know, what percentage of patients reach a certain percent threshold. One doesn't have that ability anymore because it becomes too complicated.

You know, X percent reaches 30 percent. Y percent

jam

reaches 40 percent, cetera. It gets complicated.

So I think preserving the clinically useful thresholds is very valuable, and I think we should try, but I think what we have all agreed upon is there are other bars that need to be placed there.

DR. SIMON: Does that mean you want remission?

DR. FELSON: I guess I am speaking against the motion which was I don't think we should encourage -- I think we should have the definition of major clinical improvement in there. I think it is a valuable definition.

DR. SIMON: The question here is about complete clinical response.

DR. PETRI: Let me rephrase the question because it got lost in the discussion. The question is: Should the complete clinical response and remission claims be separately described to keep the connotation of remission meaning off drugs?

I remember the controversy this morning was off what drugs. So those of you who would wish to keep these are separate definitions, separate --

DR. SIMON: Could I ask a question?

DR. PETRI: Yes, Dr. Simon.

DR. SIMON: Dr. Abramson, could you please explain to me what is wrong with saying something is good like it is

jam

a major clinical response versus a remission?

DR. ABRAMSON: No, no, no. There are two issues that I think are getting blended. I don't have a problem with coming up with a definition of remission or clinical response. I think the ACR criteria that David presented has more value to it than we have given it, that we shouldn't limit a discussion of definition major response to just picking ACR 70. We may choose to do that, but we should not lose the notion that if they are valid criteria in the new validation that they might be a way to compare drugs in different trials, sort of like a sunscreen, the SPF-5, 10 and 15. We know how much benefit you are getting from this particular drug, and that should come, perhaps, as one of the goals to help give these bench marks so we can do relative drug comparisons in different studies.

With regard to this, I have no problem with differentiating remission from clinical response, and I think the drug does make a difference in my mind, and I use prednisone as the example. I can put anybody into remission on prednisone.

DR. WOODCOCK: One comment about that -- sorry for this diversion, but your proposal is that major clinical response could be major clinical response 60 or just leave "major clinical response" out of it.

jam

The only caveat I have to this is these are comparative trials, and we are talking about a delta compared to the comparator group. Even a placebo delta will get some patients and some starting clinical states into a remission state. So it will depend on how strong the active control is as far as that delta that is achieved, and it won't be the same as oncology trials where they are basically using historical control rate for what they call a complete remission or whatever, a 50 percent response rate.

DR. PETRI: Back to the question.

DR. SIMON: What is the question?

DR. PETRI: Is it important to have two separate claims? One is complete clinical response, and the other is remission, and remission means off drugs, with an "s," off all rheumatic drugs, with an "s." That is how it is phrased here.

DR. FELSON: I would like to make the motion that it is not a valuable distinction; that I think complete clinical response and remission are essentially the same. One is defining them based on time and off and on drugs, and I don't think it is a useful distinction. I think we should just define some state of absent disease activity.

DR. PETRI: Dr. Chambers was frowning.

DR. CHAMBERS: Yes, because it has been my

jam

recollection in all of our discussions that it is off the test drug, not off all drugs. So there is a typo.

DR. PETRI: This is important because I think the panel -- remission in my view means off rheumatic drugs.

DR. JOHNSON: Yes, because otherwise you would jack up the steroids in your control arm and you have got to remission off your drugs. You can game the system otherwise. I think it has got to be off everything. You can take your antihypertensive medications.

It is an analogy to oncology. That is why we did it that way, and also, to keep it akin to the JRA situation where spontaneous remission actually occurs pretty frequently. It was to keep a parallelism there is one of the other reasons why we did it.

DR. PETRI: I'm going to bring this to a vote. We can't belabor this anymore. So, again, the question, is it important to have two separate claims, complete clinical response versus remission, where the remission claim means off all rheumatic drugs? Those of you who believe it is important to have these two separate claims, please raise your hand.

[Show of hands.]

DR. PETRI: Those who oppose, please raise your hand.

jam

[Show of hands.]

DR. PETRI: There are three opposed.

Now, we have to make a small change in the afternoon agenda. We are going to move on to the JRA section now because of people's travel plans.

I want to remind you that there is a whole page of questions directed against JRA, and that is going to be the focus of our discussion, and Dr. Patience White is going to start us out with an overview of JRA subsets.

#### **Overview of JRA and Subsets**

DR. WHITE: It is my dubious honor here to try and re-review some of the confusion that occurred yesterday, and I was asked by Lisa to address a couple of things.

[Slide.]

DR. WHITE: One is about classification, and the other is about structural disease, in particular, about X-ray progression in JRA, and then finally looking at outcome prognosis and the role that the course plays here as opposed to onsets.

[Slide.]

DR. WHITE: The classification that we are all talking about is really a classification criteria design to separate subsets, and I think that -- there are two important sets of criteria, and one is called the juvenile

jam

rheumatoid arthritis criteria. That is the one from the ACR. It is probably an American setup. The other is a juvenile chronic arthritis which is ULA. There are two of them.

The reason why I am bringing this up is that when you are looking at outcome studies, they have based these studies on these two different sets of criteria. So it just confuses this even more than where we were yesterday.

So let's sort of quickly look at this is the classification of juvenile rheumatoid arthritis, and this was validated in retrospect to cohort study of 250 patients. You can see here they do it by under 16 years. They define arthritis.

Notice that the duration of disease has to be six weeks at least, and they define in the first six months three major groups, and this is onset criteria. So you have poly-arthritis, five or more joints, oligoarthritis, less than five, and the systemic arthritis, the characteristic fever. We are talking about two weeks going to 39 degrees daily. That gives you the major areas that they are talking about.

Now, just to show you where the ULAR group is, this is their criteria. You can see that they have defined a few more groups. They still deal with the age at

jam

16, arthritis in one or more joints. They are talking about duration of disease meaning three months, and then you define your characteristic onset by these groups.

They have used the three that are used in the ACR criteria, but add that juvenile rheumatoid arthritis now is just the seropositive polies, and you have got juvenile ankylospensilitis and juvenile psoriatic arthritis. So this gives you a little bit of idea of the confusion even in the onset criteria, but I think we are basically talking about here the ACR American criteria. We are talking about onset.

[Slide.]

DR. WHITE: Let's move quickly to -- here are the subtypes, and as we have talked about them, everybody has slightly different percentages. Remember that the poly articular onset does get sort of separated out in prognosis by the presence or absence of rheumatoid factor, and by the way, they don't talk about how you measure that rheumatoid factor in the ULAR criteria. They just say its presence, and the pauciarticular group is -- there is a group that is very young in onset, high incidence of iritis.

[Slide.]

DR. WHITE: So let's now look at -- once you have got those three in your head, those major categories in terms of onset, this is some X-ray -- this is out of Jim

jam

Cassidy's text. I thought we would take his 1996 text and just show you, and if you look at the middle group here -- see this group here, advanced changes. I think that's -- oh, boy. Anyway, you can see where I am, right here, the advanced changes.

You can see in the polyarthritis group, in terms of cartilage destruction and bone destruction, you are talking about 55 to 35 percent. Even in the oligoarthritis, you are looking at some pretty destructive disease in a quarter, and in the systemic disease, it is about 50 to 20 as you can see.

So, in terms of structural disease, a large percentage of young people are going on to pretty destructive X-ray changes.

[Slide.]

DR. WHITE: Let's just look at two certain terms of the time. There are two studies. Certainly, Dan can comment on these in particular. There is a Cincinnati study that was reported by Levinson and Wallace in this Journal of Rheumatology article here. It was a Cincinnati study, and they had 114 of their 238 patients. They were getting regular X-rays.

That 114 were seen within the six months of onset. They had follow-up for a mean of 13.15 years, and they

jam

divided out by these onset type, percent of X-rays that were abnormal over that time, and then how long it took for half of them to have those kind of changes. You can see this relatively quickly. It is within two years for the poly and systemics, and a little bit longer for the pauci. So this is a pretty destructive disease, pretty quickly in a good percentage of the patients.

Now, there was another study done in Seattle, and they just asked the question, we think since the poly onset group seems to be the group that has worsening of their X-rays, they followed this group, and they had their disease for at least two years. They followed them up for 5.8 years, and you can see that two-thirds had joint space narrowing or erosions at 2.6 years. That group particularly has a rather destructive course.

[Slide.]

DR. WHITE: Finally, let's just try to think clearly about what these onset criteria are as opposed to course. This is what was coming up yesterday a little bit as we were trying to decide which group should be in trials. I think that is really the question here.

Now, the first question is: Did they all go into remission? No. This is old data, again, the same article showing you that -- dividing them up by their onset

jam

criteria. A lot of them still have active disease at 10 years.

[Slide.]

DR. WHITE: When you look at Steinbrocker, a lot of you know this, we are going to be looking at classes 3 and 4 particularly. So this is a very severely involved young person who hardly can do self-care coming out in that particular group.

[Slide.]

DR. WHITE: This is from Boll Anderson Garra's review, looking at a population base. Many of these studies here in this particular one, looking at functional outcome, are really clinic-based. So this is not a population study, and you can see that a good percentage are in Steinbrocker class 3 and 4 along the way, so that we do have -- and the question then becomes who falls into that group.

[Slide.]

DR. WHITE: I will show you the way Jim Cassidy put it together, and I am about to finish. This is why it is so confusing. So here is your onset type here, and look at the different courses you can have, and clearly, everybody will agree that there's a rheumatoid factor positive polycourse, has a poor outcome, and in most studies, we're talking about 50 to 70 percent have very

jam

severe disease, though it is a small percentage of the total of young people of arthritis.

Now, if you look at the seronegative polycourse, down at the bottom, not the ANA positive, but the seronegative polycourse of polyarthritis, it says variable, and when you look at studies, it ranges between 15 and 40 percent that have the same bad outcome as somebody who had rheumatoid factor positive.

So now you're beginning to say, well, it looks like polycourse is a problem, and that is what I am trying to get people to think about.

We are going to go to systemic disease. In systemic disease, 50 percent go into complete remission, have no problems at all. Of the remaining 50 percent, half of them that have a polycourse have a very bad course with erosions. So now you are again saying they can get rid of their systemic features, and many of them do by year -- some people feel it is up to about five years, but a lot are gone by three years, but they are left with a severe polycourse.

[Slide.]

DR. WHITE: Now, let's just look at the most common group and look at the clinical course of this particular group of the oligo. They can be mono, go on to have still under five joints, and then a group goes on to

jam

polyarthritis, and look at what happens to them. A lot of them have bone erosions and go into a poor functional class.

So, really, I am posing the question that the onset criteria might be a little confusing here. You really want to match the course and the outcome in the patients that you are going to be putting into trials. So, if you have a drug that is a lot of risks and you want to treat a severe disease, I am not sure we should be talking about onset group. We should be talking about course that are put into trials, and actually, that has been done in some of the studies.

So I am going to end there and let Lisa take over.

#### **Application of Pediatric Rule, JRA Claims**

#### **Structure, JRA Claims and JRA Drug Developments**

[Slide.]

DR. RIDER: We are going to talk now about the juvenile rheumatoid arthritis section in the document which begins on page 39 of the document, and there are really four issues for us to address this afternoon: first, the possible application of the pediatric regulation to JRA; second, the structure of claims for JRA; and then, trial design and drug development issues for JRA.

What I would like to do is present each of these topics and then present questions to the committee for a

jam

discussion after each of these.

Next, please.

[Slide.]

DR. RIDER: As you heard yesterday, the pediatric use labeling regulation was adopted by the agency in order to facilitate labeling of agents for use in pediatric populations. The regulation states that when the course of the disease and the drug's effects are sufficiently similar in the pediatric and adult populations to permit extrapolation of adult efficacy data, then pharmacokinetic, pharmacodynamic and safety studies are required for pediatric labeling of the agent.

This regulation applies to new applications received, as well as retroactive applications and licensed products.

[Slide.]

DR. RIDER: In considering whether the pediatric rule applies to JRA, we first need to consider the current realities and difficulties in drug development for JRA.

First, we have only three drugs approved for use in JRA. Yet, we have widespread off-label use of drugs occurring in this patient population.

JRA is also a very rare disease, and only 3 to 5 percent of patients with RA have onset during childhood.

jam

Furthermore, it is difficult to obtain controlled clinical trial data in pediatric patients in general, and also, of necessity, because of the rarity of the disease, JRA efficacy trials are multi-center.

Why this concern for JRA at the FDA, our perspective is that the drugs are being used now off-label; that there are many obstacles to getting definitive trials done; and that applying the pediatric regulation will promote safer use of these agents in these children.

[Slide.]

DR. RIDER: We need to consider whether the pediatric rule applies to JRA. As you hear from Dr. White, only about 5 percent of the patients have rheumatoid factor positive of polyarticular JRA, and those patients, by our criteria, are considered the same as adult RA patients, but it is our observation that, really, patients with a polyarticular course, as Dr. White pointed out, both seronegative and rheumatoid factor positive patients are really similar to adult RA in a number of ways.

First, they share a common immunogenetics that many of these patients, but not all of them, have common Class II and reach the genetic risk factors, including shared RA epitopes.

Secondly, their immunopathogenesis is very similar

jam

to adult RA, including what we know about cytokines in the joint, lymphocytes upsets, and inflammatory markers in the joint of these patients.

As you heard from Dr. White, the disease course is often erosive and destructive, and there is often high morbidity, and these patients to date have had similar responses to therapy.

[Slide.]

DR. RIDER: In the rheumatology working groups proposal for applying the pediatric rule to JRA, our proposal is that the pediatric rule would be applied to the signs and symptoms claim only; that efficacy studies would be required in JRA for all other license or claims.

We plan that extrapolation of adult RA efficacy data would be for polyarticular JRA patients to clarify rheumatoid factor positive and seronegative polyarticular patients with a polyarticular course.

This would be when there is only biologic plausibility that the agent would have a similar effect in JRA as in adult RA.

Analogous to the application of the pediatric rule to other pediatric populations, additional pediatric dosing and safety evaluations would be needed in polyarticular JRA patients to obtain a label for polyarticular JRA.

jam

[Slide.]

DR. RIDER: Our three questions to the committee, to each be considered independently, are as follows. Does the committee agree with the proposal to use the pediatric rule to grant labeling now for the signs and symptoms claim without any further data for all JRA subsets for currently licensed NSAIDs, methotrexate, and prednisone, based upon published controlled trials in JRA and their general use?

[Slide.]

DR. RIDER: Second, does the committee agree with the proposal to use the pediatric rule to provide eligibility for labeling for polyarticular JRA only for certain other agents currently licensed for adult RA, but not licensed for JRA? These currently include auranofin, gold sodium thiomalate, hydroxychloroquine, and penicillamine. We would expect that supporting pediatric dosing and safety data would be needed to support a label for polyarticular JRA.

[Slide.]

DR. RIDER: Third, for experimental agents for adult RA, does the committee agree with the proposal to apply the pediatric rule for polyarticular JRA when there is biologic plausibility that the agent would have a similar effect in adult RA?

jam

We would also like to ask about the extent of dosing and safety data that would generally be needed in applying the rule to JRA and whether these data need to come from polyarticular JAR patients. It has been the prospective of the agency that each patient population has different safety and dosing issues.

DR. PETRI: Thank you.

Now, for this discussion, I would like to have our pediatric rheumatologist be the most active participants, and I would like to invite the pediatric rheumatologist in the audience, including Dr. Silverman, to participate as well.

So let's address the first question which his, right now, are there drugs that should receive symptoms and signs claims, and the ones suggested were the NSAIDs, methotrexate, and prednisone.

Dr. White, why don't we start with you.

DR. WHITE: Shall I answer yes to the question?

DR. PETRI: Well, we would like you to share your wisdom with the committee. If that is a yes, that is fine.

There is a three-paged handout from this morning. I am just rephrasing the questions that were shown on the overhead.

Dr. White can go ahead and start us off here.

jam

DR. WHITE: Well, you know, I think that we are using all these drugs now in kids, and there is a lot of data out there using them. I mean, some of the nonsteroidals have been looked at in children. In collaborative drug study, methotrexate clearly has been looked at. So my sense is the answer is we have the data and go ahead. It has been used in all subsets. There is clinical experience here.

DR. PETRI: Let me ask Dr. Barron.

DR. BARRON: I agree. We have been using these drugs for a long time, and as was said, we have the data on them, especially for these categories of drugs, NSAIDs, methotrexate, and prednisone. In fact, we have been using other drugs and have data on those as well. So, at least for this question, I would say yes.

DR. PETRI: Dr. Silverman? Dr. Lovell? Any additional comments?

DR. SILVERMAN: I think, again, this is an easy question. I think those three drugs or two drugs plus one class of drugs are commonly used and have been shown to be effective. So that is an easy yes for those.

If I can just comment, though, since I am standing here, anyway, Dr. White's definitions of subsets and how they really have to be used, not the ACR definitions, I

jam

think we have to maybe, as we get to the next question -- it could be just reworded to say -- or one of the other questions -- polyarticular course of JRA, and that may be very useful.

DR. PETRI: I agree. I think that was the most important point that Dr. White made that it's not the onset that is going to be important in clinical management. It is the course.

Now, let me open up this first question again for the entire panel. Dr. Fernandez-Madrid?

DR. SILVERMAN: I just wonder if that is necessary as opposed to just saying in JRA. Whether one in the panel would be happy with saying that methotrexate has an indication for pauciarticular, I just wonder about the use and the efficacy of safety data in pauci rather than just saying JRA and leaving the subset definition out. That would be my preference personally.

DR. PETRI: Thank you.

Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I would be in favor of answering yes to these questions, particularly for methotrexate and prednisone for the reasons that you gave.

I have a question. Is there not a problem with the use of aspirin and sometimes indocin in children?

jam

DR. PETRI: Did you want to comment on that?

DR. BARRON: I think that aspirin and indocin are used by a small number of rheumatologist, and in fact, most of the other nonsteroidals are used instead.

In all of the pediatric diseases, at least aspirin is used in Kawasaki and rheumatic fever and is rarely used in JRA anymore.

DR. WHITE: Remember, indocin is used in babies to close patent ductus. I mean, there is a lot of use of these drugs.

DR. LOVELL: Indocin has a definite role, and I think its tolerance in the pediatric population is as better or comparable to the adult RA population. So I don't think, clinically, it is much different in its side effect profile in children and adult.

I have a question for the FDA people. If we vote to say yes to the wording of this particular question, what does that mean for new NSAIDs that come down the pike? Nothing?

DR. PETRI: Dr. Chambers?

DR. CHAMBERS: It depends on how closely we believe that the new NSAID is the same as what is the meaning of this class as it is currently approved.

If we think that it is essentially the same, we

jam

would go ahead and apply this there. If there are reasons to believe that it is different based on claims that are made in RA that it is different, we would not apply it directly as part of what is written here.

DR. LOVELL: One question or concern I have is that the methotrexate was really studied in polyarticular course JRA, and I would be a little reluctant, I think, to license its use for pauciarticular JRA.

I think if our licensing somehow drives the labeling and the package insert information, I think we need to be careful how we kind of summarily pass these drugs along.

NSAIDs are used in all three subtypes. Methotrexate is used in systemics and polys, but I'm hesitant to kind of give a blanket approval for its use in paucies.

DR. PETRI: Dr. Simon, you had a comment?

DR. SIMON: I have a question.

DR. PETRI: Let me ask Dr. White. Would you agree that we should have a limitation on methotrexate?

DR. WHITE: Well, you know, I think, really, we should talk about course, and it is going to come up in the next one, too. I mean, I think if pauci onset has a poly course, they are going to get on these drugs.

jam

DR. PETRI: Well, but, then, Dr. Silverman suggested that we use the term "JRA" in this. Do you want us to go back and have our motion in the form of polyarticular course, even for this first question?

DR. RIDER: I am aware from various academic centers that it is used in the few pauci patients --

DR. PETRI: I know it is.

DR. RIDER: -- and having similar response effects.

DR. PETRI: I think we are ready to phrase the question.

DR. SIMON: Michelle?

DR. PETRI: Yes, Dr. Simon.

DR. SIMON: I just want to ask two questions about methotrexate as we understand its use in children before we take a vote.

The first is that we know in cancer therapy that there are significant effects on bone, both osteoblast function, as well as the inducement of osteoporosis in regression fractures.

Certainly, understanding the glucocorticoids can cause lots of problems as well. Has this actually been well studied in the use of methotrexate in children for this particular indication, at least as yet, and are the

jam

pediatricians concerned about that?

The second one is the data about Cox-2 knockout and what we now know about the developmental problems associated with mice in Cox-2 knockout, and that we have not had to deal with much of that in kids growing and using drugs that are highly specific in knocking out something that may be very important developmentally, at least in animals.

Is it possible that those would then be considered nonsteroidals, and so this kind of message would not be the appropriate one to give to potential sponsors out there in the development of these drugs?

DR. LOVELL: I can answer your first question about methotrexate on the bone in JRA patients. First, it has not been well studied, but we did look at it in our bone study to see if methotrexate, per se, was a predictor of poor bone mineralization, and it wasn't statistically associated with the people who had normal versus low bone mineral density.

The studies in pediatric oncology patients were utilizing much higher doses of methotrexate, but I think the data we have from even the low does of methotrexate being inhibitive of osteoblastic function in vitro is of concern, but we haven't seen now over a decade of significant

jam

clinical use of this product an increase in the risk of compression fracture.

DR. WHITE: The second question, I think, they would fall into a different category completely, wouldn't it, Dr. Chambers? That would be my understanding. This isn't the same class of drugs.

DR. CHAMBERS: To the extent that the Cox-2 selective agents differentiate themselves from other NSAIDs, the farther they get to differentiating themselves, the farther they would get to not having themselves included in this proposal.

DR. SIMON: A very safe answer.

DR. CHAMBERS: At this point, I do not have sufficient information on Cox-2 products to be able to give you a definitive answer. At the point that we get one approved, I will be able to talk to you much better on it.

DR. PETRI: I think that the committee's clear message here, though, is that we don't want our vote to be generalized in the absence of data.

DR. JOHNSON: Could I ask a question of the pediatric rheumatologist, one or two of them? Is it believed as strongly that methotrexate works in paucies as it is believed that nonsteroidals or steroids work in paucies? Do you follow me? Did you understand the

jam

question?

DR. WHITE: I mean, I don't particularly use it in paucies.

DR. JOHNSON: Because the kids aren't sick enough?

DR. WHITE: A joint injection with steroids works, and so it hasn't come up, although I have had patients referred to me on it, and I have taken a lot of patients off.

So I am trying to answer your question, and I am not sure I can.

DR. JOHNSON: I think part of the rule is that if there is biological extrapolability, which maybe there isn't with paucies --

DR. WHITE: Right.

DR. JOHNSON: -- but if there is, then you don't need trials in a sense. I mean, you may need some safety of PK, but beyond that, you are relying on your experience where you believe it works and other people's experience that believe it works, too.

DR. LOVELL: Well, my experience there are pauciarticular course patients in whom the joints that are involved, for example, their hips and their knees, they are significantly disabled by it, or hips and ankles or knees and ankles, and methotrexate in those patients seems to work

jam

as effective as it does in patients with more joints.

It is not a question that the response is in any way different. It is just a question of is the disease severity enough in that particular patient to warrant it, but I think the drug efficacy and mechanism of action would be entirely the same in those patients.

DR. PETRI: Dr. Abramson?

DR. ABRAMSON: Michelle, I just had a question about the language of the question that we are voting on because I had some concerns about what we are actually being asked of vote on.

DR. PETRI: May I rephrase it?

DR. ABRAMSON: Well, I just want to raise the issue of -- we are basically hearing that this is the standard of care among our best pediatric rheumatologist, and that I certainly have no problem accepting.

I have a problem for reasons of not accepting that all these subsets are really just young kids with rheumatoid arthritis using the pediatric rule as written in this question No. 1 as the reason that I support the view that these drugs may have appropriate use and support by this committee for use of pediatric population.

So I think there are two separate issues here that are bundled into the way this question is written, and it is

jam

going to be more relevant, I think, for the discussion of the second question.

DR. PETRI: I wanted to rephrase the question. So let me rephrase it now and let's get to a vote.

There is immediate reason to give the symptoms and signs claims to NSAIDs, methotrexate, and prednisone for a JRA with a polyarticular course. I am willing to have a second vote later, but the first vote will be on polyarticular course.

Those who are in favor, please raise your hands.

[Show of hands.]

DR. PETRI: Those opposed?

[No response.]

DR. PETRI: I see no opposition.

Now I am going to have a second question on whether methotrexate should have a symptoms and signs claim for all JRA including pauciarticular.

Those who are in favor, please raise your hands. The question is should methotrexate have a symptoms and signs claim for all JRA, including pauciarticular. All those in favor, please raise your hand.

[Show of one hand.]

DR. PETRI: One in favor.

Those opposed?

jam

[Show of hands.]

DR. PETRI: So the consensus is against that, but obviously, I think this could be reconsidered when there are more data.

DR. LOVELL: Should we reconsider if we said as polyarticular?

DR. PETRI: We are talking about course now, if the child has a pauciarticular course. There wasn't a consensus that we had enough information, and in fact, Dr. White was suggesting that she didn't think it was the most appropriate clinical management for a pauciarticular course.

DR. WHITE: Well, it gets very individual. I think that is the key. If the four joints are both hips and both knees, that is a very different thing than if the four joints are a wrist and two fingers, and I think that is the point that Dan was making. Am I correct?

DR. LOVELL: Right, yes. I am getting confused by the questions here about methotrexate. Can you reiterate for me what we have just done here?

DR. PETRI: The first question was on polyarticular course.

DR. LOVELL: For all three?

DR. PETRI: For all three.

DR. WHITE: Right.

jam

DR. PETRI: The second question --

DR. LOVELL: Okay. So we said yea to that?

DR. PETRI: Yes.

DR. WHITE: Yes.

DR. PETRI: The first one was positive. The second one, we just talked about, methotrexate and pauci, that was not.

Now, there is a third question which is: For indocin and prednisone, is there enough current information to justify a symptoms and signs claim for all JRA including pauciarticular?

Those who are in agreement, may I please see a show of hands?

[Show of hands.]

DR. PETRI: Those opposed?

[No response.]

DR. PETRI: That passed.

Can anyone think of a permutation I have left out here?

Now, the second question is: Can we invoke the pediatric rule to allow eligibility of the following drugs, auranofin, gold hydroxychloroquine, and diphenylamine, for a polyarticular course JRA, obviously with the proviso that there would have to be dosing and safety information?

jam

Is there a discussion? Again, I would like the pediatric rheumatologist to lead this discussion.

Dr. Lovell?

DR. LOVELL: I can address that issue for three of the four drugs quite directly. We have done prospective placebo-controlled multi-center trials of auranofin, hydroxychloroquine, and diphenylamine in polyarticular course JRA, which is the set we are talking about, and none of those drugs were different than placebo. So I would be a little reluctant to reverse that significant effort.

DR. JOHNSON: Unless your assay failed, but three failures in a row, I guess, would be unlikely.

DR. LOVELL: Well, I think that is the trial data that we have.

DR. RIDER: How about your placebo response rate above 40 percent or 50 percent in those trials?

DR. LOVELL: For the hydroxychloroquine and diphenylamine studies, the placebo response rate was about 40 percent for auranofin. Yes, for those three drugs, it is about the same, but on the other hand, they were placebo-controlled studies utilizing our best clinical measures, and in one of those three drugs was there demonstrated efficacy above placebo.

DR. PETRI: Really, what you are suggesting is

jam

that this pediatric rule falls apart for some reason. Our biological plausibility went down the drain.

Dr. Silverman?

DR. SILVERMAN: It brings up a very interesting point. If those drugs came on the market now, I would support the pediatric rule to be applied. However, because I know they don't work, I can't support it. So it is an interesting twist here. So, if you have a drug that is on the market and we have done the trials, which also back the clinical impression that these three drugs do not work, therefore, I have difficulty saying we should use a pediatric rule.

However, if diphenylamine came on the market today and it was shown to work in adult RA, we would apply the pediatric rule as was argued yesterday. So it is an interesting conundrum, and there is no way out of this.

DR. LOVELL: Well, actually, I think there is a way out of it, and I think part of the biologically plausible information we have about polyarticular JRA is that the drug we have to date that has been efficacious has been methotrexate. So, when we get into these discussions about application of a pediatric rule, we say a part of the biologic plausibility about the efficacy of this drug is that it needs to be in the same general category of efficacy

jam

as the only second-line agent that has ever been shown to be efficacious in JRA, which is methotrexate, in a very strong sense.

So that would be how I would get out of this conundrum is to interpret the biologic plausibility wording of this pediatric rule to say that these drugs that are tried in adults that we are thinking about downloading into pediatric rheumatology via the pediatric rule should have efficacy that is in the same general category as methotrexate.

DR. RIDER: So, then, is the pediatric rule only applicable if RA patients are studied with a methotrexate background therapy?

DR. LOVELL: I would caution that we probably ought to because the last thing we want to do is kind of put into our pediatric regimen treatments that are going to be inefficacious, and the pediatric rule, if we applied it at this point in time, three, perhaps four drugs that aren't efficacious because I don't know about gold shots. WE have never studied it.

MR. STRAUSS: I work for pharmacy at Upjohn, the clinical development.

Sulfasalazine is approved for adult RA in the U.S. as of last year, firstly, as a point of information and it

jam

is left off the list.

Secondly, there was a study that was presented at the ACR last year from Fonderossum and the Holland group in which they, in fact, tested sulfasalazine against placebo in JRA and showed it a positive result there.

DR. LOVELL: I think the offer still exists for companies to come in and test drugs in JRA in the regular way. It is just we have to be careful we don't kind of do damage to our clinical care practices by over-application of the pediatric rule.

There is nothing to prevent sulfasalazine, for example, or other drugs to try to get approval and indication for signs and symptoms in JRA in the more rigorous way.

DR. PETRI: Dr. Strand?

DR. STRAND: Well, I am just a little curious whether you might want to be this restrictive with yourself because, in fact, you may have drugs that so far in placebo-controlled trials haven't been efficacious, but perhaps a combination of some of them would be.

If they are not in current use, I can see that perhaps you don't want to extend the label, but it doesn't make sense to me that we should, by definition, take these three particular products and then turn around and say,

jam

well, then there is no reason to apply the pediatric rule because, from the other side of it, you have been arguing that you don't get the opportunity to have products to test in JRA or to even use. It is such a small indication, it is difficult to get sponsors interested in the products.

So I am not quite understanding why you want to restrict yourself so much, even as much to say that pauciarticular with methotrexate because you have got a safety profile in JRA now from the published data. If somebody wants to use it or if perhaps methotrexate plus sulfasalazine, plus hydroxychloric may turn out to be effective in JRA. You are trying to get labeling where there is at least some evidence of benefit without undue risk, and you then have the opportunity to use it in practice. Otherwise, with our current medical situation, you won't be able to use any of these products.

DR. PETRI: Dr. Silverman?

DR. SILVERMAN: I think Dr. Lovell answered my conundrum by, one, saying that to date, if you show -- I don't think it has been the background of methotrexate if certainly there is a study showing against methotrexate would be adequate. I think showing it in the same class would be certainly the plausible explanation, and then a new class which shows to be as effective as methotrexate. It

jam

would also be pretty convincing data to apply the pediatric rule, and I think the other question, it is hard. You cannot undo what we know didn't work in proper-controlled studies, and I would emphasize, then, if we wanted the other studies, they should be done, but not using the pediatric rule.

DR. PETRI: Yes, Dr. Barron.

DR. BARRON: I have a question for Dan. Are you suggesting that we use methotrexate as the gold standard for pauciarticular JRA?

DR. LOVELL: No, no, no, not for pauci because the pediatric rule doesn't apply to pauci. It applies to poly, and I think we are asking that they change the word to "polyarticular course JRA." So it is the pediatric rule, as we have been told about, that applies only to polyarticular JRA. So it wouldn't be the gold standard for pauci. It wouldn't have much relevance at all.

It is just a mechanism by which we can get drugs labeled for JRA, and I think we ought to be careful we won't make the mechanism so facile that we get drugs labeled for JRA that are not efficacious.

DR. PETRI: Let me try to rephrase the question. The question was whether using the pediatric rule, the following drug should be eligible for symptoms and signs

jam

claim, the auranofin, gold hydroxychloroquine, and diphenylamine, with the understanding there would have to be dosing and safety data, and whether our discussion has focused on the fact that, although the pediatric rule might apply, the drugs have not been shown to be efficacious.

Those who would like to have these drugs made eligible under the pediatric rule, would you please vote yes.

[No response.]

DR. PETRI: Those who are opposed?

[Show of hands.]

DR. PETRI: So the consensus is no.

DR. JOHNSON: Can I make a comment now that the vote is over?

DR. PETRI: Yes, Dr. Johnson.

DR. JOHNSON: I really don't think we are out of the conundrum, as a matter of fact, and I am a little worried, too, that we are setting the hurdle too high for kids.

I mean, it may well be that there was a structural reason why those other three trials all failed that had nothing to do with the drugs, and maybe it is just too much variability or whatever.

If it is something structural and we don't know

jam

what it is, you might expect it to have plagued all three trials. Let's say each of those trials was 10 times the number of patients. Maybe it would have showed the effect. It is just a smaller-than-methotrexate effect, but as it stands now, nothing that doesn't match methotrexate is going to even get considered for the rule. I mean, that would be the logical conclusion from what you are saying.

DR. PETRI: Well, aren't we almost saying that about adult RA, too?

DR. JOHNSON: No. You could still do a -- no, you are not.

DR. PETRI: Well, we said with a preferred active control, it would be methotrexate in the adult RA studies.

DR. JOHNSON: Yes, but you can still do three-month placebo-controlled trials and prove all kinds of things that are milder than methotrexate and get them approved, if you want them.

DR. PETRI: Yes.

DR. LOVELL: In answer to your question, I am perfectly comfortable with that because it is not the only alternative for people to get an JRA approval for signs and symptoms. It is just the kind of easy way, and I want to make sure that the ones that come through by the easy way are truly going to be drugs that are going to benefit the

jam

patients to whom they are given, but if you have a sulfasalazine, for example, that may not be as efficacious as methotrexate. Then, you can still look against placebo and see if it is beneficial and still get an indication.

The other thing is that the same kind of conditions and methods and outcome measures were used in a methotrexate study, and it did work. So it is not kind of a fatal systems there.

DR. JOHNSON: Well, it may be fatal to mild drugs and successful for moderately active drugs like methotrexate. I mean, that is what I am alleging.

DR. PETRI: Let's move on to the third question which is for experimental drugs. Can the pediatric rule be invoked for the polyarticular course JRA if there is biologic plausibility and there is dosing and safety data? Any discussion before we bring that to a vote?

Dr. Abramson?

DR. ABRAMSON: Yes. I guess I would just reiterate what I was alluding to before that I personally have a serious problem with the pediatric rule, per se, in these diseases. I think it is not intellectually honest to think that these diseases, because there are cytokines and cells present in the joints, are similar enough for us to give carte blanche in this kind of notion. I think each

jam

disease has to be studied separately.

I am sensitive to the pediatric issues, but I think the votes that we took on the last two questions illustrated that when the pediatric rheumatology community has a sense that drugs are effective in the presence or absence of trials, there is some compelling support among this kind of committee to endorse those drugs, but I think to say there is something called a pediatric rule and then to say polyarticular course and even break down the notions of these somewhat arbitrary diagnoses of JRA, it is not intellectually attractive enough to sort of cover over this kind of vote.

DR. PETRI: Dr. Simon?

DR. SIMON: I would like to expand on that just a little bit more. In fact, you know, if the standard of care in the community is such, I am hard-pressed to suggest that I know more about that than someone else does. We are now in experimental therapy, not talking about standard of care, and I am very uncomfortable, particularly as it relates to -- I have an inherent skepticism to believe that you can translate the biology of these diseases, just because we've named it polyarticular.

So I defer to my colleagues who are pediatricians and who know how to take care of these patients and say to

jam

them I believe that you are doing a good thing and this is the right thing to do, but to then border into the experimental, I get a little uncomfortable with.

DR. PETRI: Let's hear from the pediatric rheumatologist. Dr. White?

DR. WHITE: We are all in the same conundrum here.

There is the scientific view that no, you would like to put them all through trials just the way we have been talking about, and then there is the practical view that the chance that we will be able to do that is very small.

DR. LIANG: But you are making that probability --

DR. WHITE: I know.

DR. LIANG: -- even smaller by giving them an easy label. No one is going to sink any money into funding a multi-center trial, then, or even the toxicity.

DR. WHITE: I know.

DR. LIANG: I think that is far more important.

DR. STRAND: I would like to respond that they will.

DR. LIANG: They haven't.

DR. STRAND: They will, they will, and I think that, in fact, the pediatric rule is an incentive for there to be PK in safety studies. That they haven't to date

jam

doesn't mean that they aren't doing or going to do it and you don't know about it yet because it hasn't come for approval.

I think it is an important thing to think about because there is really very little other way to get supporter interest in JRA trials.

DR. LIANG: I don't see now you can reassure us in that regard.

DR. STRAND: Why not? If I have a client that is doing it, that is not a reassurance?

DR. LIANG: One client?

DR. STRAND: Well, no. I mean, I can't say that --

DR. LIANG: I mean, I think the action speaks louder than words. There has been no major support.

DR. STRAND: No. I just said I can't say how many, but yes.

DR. PETRI: I think what the committee is saying is that we don't feel comfortable with a blanket approval of the pediatric rule and that each drug is going to have to be considered on an individual basis. So, if I could bring that to a vote, those of you who agree with that consensus statement, please --

DR. WHITE: Wait a minute. You would apply it to

jam

individual drugs as opposed to a blanket. Is that what you are saying?

DR. PETRI: Yes, exactly, that we cannot have a blanket pediatric rule for all experimental drugs tested in adults; that each drug has to have the pediatric role and vote individually on that drug's merits.

Dr. Barron?

DR. BARRON: I think that we just need to also comment that as each drug is considered, we need to weigh the risk and the benefits, and that there are certain categories of drugs that are going to have far more risk than benefits, while other categories of drugs will have more benefit than risk. So I think each drug will be best to be considered individually.

DR. PETRI: Yes.

Now, those of you who agree with this statement, please raise your hands.

DR. JOHNSON: You mean the statement you just made.

DR. PETRI: Yes, yes. I got tired of rephrasing it.

[Show of hands.]

DR. PETRI: And those opposed?

[No response.]

jam

DR. PETRI: So there is complete agreement among the committee.

DR. RIDER: Could I clarify that? Do you mean for us to bring forward each drug to this committee for your consensus whether we could apply the pediatric rule?

DR. PETRI: I think that is an agency decision, actually. Obviously, this committee would be very happy to discuss individual drugs, but I don't think that you should feel that all drugs have to come to us for that decision.

As we move back to the afternoon agenda, are there any issues on the preclinical and early clinical sections of the guidelines?

Excuse me. Dr. Rider?

DR. RIDER: We are going to move on now to the JRA claims structure and questions regarding that. The JRA claims structure was set up to be parallel to the structure for adult RA, to include clinical signs and symptoms, function and quality of life, prevention of structural damage, complete clinical response or remission, and major clinical response.

Several of these claims are still undergoing validation. Methodologies are still being developed, and for complete clinical response or remission, we reached a consensus definition that our JRA workshop in July from 13

jam

pediatric rheumatologist who attended the workshop.

Our questions regarding JRA claims structure --  
next overhead, please.

[Slide.]

DR. RIDER: First, what would be the appropriate trial durations for JRA claims? The clinical signs and symptoms, three months is proposed with biologic trials of at least six months. For function or quality of life, while a 12-month time point had been in the document for adult studies, we had proposed a shorter time for pediatric studies given that a six-month time span is relatively long in the life of a young child.

Structural damage and complete clinical response or remission trials were proposed to be for one-year duration.

Second, should the trial durations change if the drug is already approved for the desired claim in adult RA?

[Slide.]

DR. RIDER: Third, are there existing or emerging databases to define major clinical response for JRA, as Dr. Felson is going for adult RA?

DR. PETRI: I don't think we want to repeat the whole discussion we had about adults. So let's try to focus this.

jam

Let me ask Dr. White for her comments about those questions.

DR. WHITE: We were sort of laughing about the issue of kids make more strides than adults in terms of functional and quality of life in a faster manner.

I don't know if this is arbitrary to me. I think it is reasonable. So I don't know. We debated this in the adults, and I think the same debate occurs in pediatrics. I think that is really what the issue is. We found it hard in adults. We are going to find it equally hard in kids, and we sort of settled on things. I think that we could probably settle on this group as they stand.

DR. PETRI: Dr. Felson?

DR. FELSON: Two comments. One is a short one, and another is longer.

First, the answer to the third question, existing or emerging databases, at the end, Dan could comment on this better than I. I think the database on JRA trials is the best there is. It has got all the cooperating trials data. Maybe FDA has more data than that, but Ed certainly marshaled it to look at these questions, and I would ask him to do the same things we just did for adult RA.

He has got less data on the efficacious drugs than we do because there is only a big methotrexate trial, I

jam

think, but that would at least give you some feeling for what the response rates are going to be, which I think you need.

Then, I think we come to the more substantive comment. We come back to the time issue that has plagued us all day. The first claim structure questions are a series of claim that relate to time. To be honest with you, it isn't clear whether they relate to time of response or time of trial.

I think that perhaps needs to be specified, but my suggestion would be to just get rid of all the time. I don't know why the time has to be here at all. I think what you are interested in is that JRA produces these -- that treatment of JRA produces these improvements, and then a particular treatment can be characterized with respect to what time -- how long that improvement might be or whatever.

I don't think it is necessarily in the crux of the matter as to whether the drug is efficacious. I think that all the time constraints here -- structural damage is one where I think that might be an exception because, generally speaking, it takes a year to show change in structural damage, but even that is not something the FDA necessarily needs to mandate.

DR. JOHNSON: So a one-week remission has

jam

credibility? I mean, that is the issue.

DR. PETRI: This is the same debate we had this morning.

DR. FELSON: Do you want to have a symptoms and signs claim for a nonsteroidal in JRA? I think the answer is yes. You don't need a three-month trial for that. You can have a one-week trial for that, a two-week. What is the problem with that?

DR. PETRI: The problem, again, is we are going to need some safety data, too. That was the crux of the discussion this morning.

DR. FELSON: Then make it clear that it is not an efficacy issue that you are asking for to find data on. It is safety.

DR. JOHNSON: No. You know the durability of a nonsteroidal. That is why you don't have to bother with it, and you can just do a one-month trial possibly, but if you are talking about new interesting agents, we have to think about the time duration. We can't ignore it because it bears on regimen and durability.

DR. CHAMBERS: It is this benefit to risk, and if you don't study for some duration along there, you have no chance of finding out what any of those risks are, and you can't make the evaluation.

jam

DR. PETRI: Dr. Lovell?

DR. LOVELL: From a pediatric point of view, I don't have any problem with the time limitations set in here except with the distinction of maybe this morning for biologics to try to make a distinction between duration of study and duration of clinical response because of the novel way some of these agents work and that sort of thing.

I don't think pediatric rheumatology has been in any way disserved by the, I think, current requirement to do three-month trials, a minimum of three-month trials for nonsteroidals, correct? I think that has not been the disservice to pediatric rheumatology, and the rest of these kind of time claims have some base validity, if you would, especially the one-year for structural damage and remission. It has to have some kind of durability to it, and one year doesn't seem like a bad fix.

So I don't have any strong reservations about approving these current time limitations with the caveat about biologics.

DR. PETRI: Dr. Silverman?

DR. SILVERMAN: May I just make one comment about the structural changes? I have a little problem with the one-year, and that is because of the delay in X-ray changes appearing because of the amount of cartilage and unossified

jam

bone.

So I think that if we use X-ray as the structural change, we may miss either -- well, progression structural damage just because of the insensitivity of this technique in a young child.

So I think when we use X-ray as a gold standard, one year is too short to get a structural damage claim, but if we went to MR or some other thing, we would actually denote structure. Then I would have no problem with one year, but I do with X-ray as the gold standard.

DR. JOHNSON: It is a minimum of one year, and if a drug actually shows it in one year by X-ray, you wouldn't dispute that.

DR. SILVERMAN: Yes, I would because I am saying it may not appear for two years.

DR. JOHNSON: No, but if it does appear in one year, you can't deny it.

DR. SILVERMAN: Then it has damage. I saying you can never show it does not have structural damage within one year. That is all I am saying, in a young child.

DR. PETRI: I believe the consensus is that, except perhaps for structural damage, the time suggested for the different claims are reasonable. Is there any dissention to that?

jam

[No response.]

DR. PETRI: I think the second question was whether should the trial duration change in the pediatric population based on adult data. Is that correct, Dr. Rider? That was your second question.

Is there any discussion about that?

[No response.]

DR. PETRI: I can't see any reason to shorten the trials.

DR. CHAMBERS: Can I just ask for a clarification? We are talking about the claims now would be the same as how we had modified them in adult?

DR. PETRI: You better rephrase that.

DR. CHAMBERS: We talked about the quality of life and a number of things being changed when we were talking about adult. We are applying -- the way we ended up with adult would not be mimicked in pediatric?

DR. PETRI: No. Basically, it is as presented to us for a pediatrics. We are not trying to generalize the duration of trials from adults, but what was presented to us by Dr. Rider appeared to be reasonable except perhaps for the structural claim, the duration should be longer than 12 months.

Dr. Felson?

jam

DR. FELSON: Since I can't convince you all to discard the time, let me try to convince you to modify it because you are making it very hard. These are hard enough trials to do with few enough kids. I think demanding real long trials for some of these outcomes is -- and by long, I mean a year -- is very tough.

I would suggest the complete and major clinical reasons and remission be the same as the clinical symptoms and signs. Remember, we got to distinguish here between duration of trial and duration of response. I think that is what everyone is talking about.

So I would suggest that we try a three- to six-month trial duration for a completed and major clinical response and remission if we are going to demand the same of clinical signs and symptoms, and perhaps within that, demand that there be some durability of the response.

I don't think it is necessarily reasonable to ask for a one-year-long JRA trial that just looks at clinical response.

DR. PETRI: Yes, Dr. Siegel.

DR. SIEGEL: I would like to clarify the second question a little bit. This is asking whether trial duration should change if the drug is already approved for the same claim in adult RA.

jam

The question here is that in the first question for signs and symptoms, the trial duration is suggested to be three months or six months for biologics, and if a biologic has already been shown to be effective in adult RA in six-month trials and that the beneficial effects do not wane at six months, the question is would it be possible to do a three-month trial with a biologic in JRA.

The reason for asking this question is that the agency is very interested in trying to get efficacy trials done in JRA with some success, and I think our ability to succeed may depend on whether a three-month trial would be sufficient.

DR. PETRI: I think that is better addressed study by study. I don't see how we can give a blanket recommendation that you need a shorter time period in a JRA trial. Let me ask if there is anyone on the committee who disagrees.

I think we feel very uncomfortable, the same way we felt uncomfortable with a blanket pediatric rule because we have seen how that can fall apart.

DR. JOHNSON: But what he is saying, if you didn't have the discomfort from worrying about the durability in adults and you have already got the data, two trials, let's say, that show that it is effective at six months, is it

jam

that big of a leap to presume that a three-month study is --  
is it too much of a leap to accept three-month data in kids?

DR. PETRI: We are saying the same thing, aren't  
we, does the pediatric rule always apply, is there biologic  
plausibility, and there is not.

DR. JOHNSON: This has nothing to do with the  
pediatric rule.

DR. PETRI: No, it was the same argument, Kent,  
because we saw a lot of drugs that worked in adults, and  
they didn't work in the kids.

So I think if we are going to do this, we ought to  
do it right and make sure it worked in the kids, but that is  
my own opinion and the committee should chime in here. I  
can't see trying to shorten that process.

Again, it should go, I think, study by study, drug  
by drug.

Dr. Lovell, do you want to comment?

DR. LOVELL: No, I agree completely with what you  
just said.

DR. CHAMBERS: I am now a little confused. When  
we were saying that the biologics -- we didn't like the six  
months for the biologics, is that because we want the drugs  
to be six months or we want the biologics to be three  
months?

jam

DR. PETRI: No. The issue was a biologic has been shown to have symptoms and signs claim in adults in the six-month trial.

DR. CHAMBERS: No, I am backing up. Originally, which way did we have discomfort that there was a separation between drugs and biologics? Which way did we fix it?

DR. LOVELL: The discussion we had this morning about biologics was that we made it six months for biologics because we wanted to allow a time to observe for the development of antibodies, and I think that is a very reasonable idea, but on the other hand, three months of clinical benefit from a biologic that may be only administered in a one-time injection, that might be sufficient clinical benefit for that particular biologic. So we shouldn't get confused by saying we are requiring six months of clinical benefit on a blanket statement for biologic agents. It is just that we need to have six months of study duration to give us time to observe for unusual events.

DR. SCHWIETERMAN: That is right. The need in biologics is to characterize the efficacy outcome because of concerns about delayed onset of immunogenicity and so forth, but three months would be adequate to measure that, so long as there was a characterization of the entire six-month

jam

course.

DR. CHAMBERS: But there is not a separation between drugs and biologics. You have a drug that also takes that long. You follow with that. You are just saying there is always a minimum of three months, and it may be longer than that if you need that to characterize.

DR. PETRI: I think we would like to move on, unless there is a question about preclinical and early clinical aspects of the guidelines. It would be equivalency trials, and Dr. Chambers is going to discuss equivalency trials.

DR. CHAMBERS: You are actually supposed to give me the preclinical so I have time to set this up.

DR. PETRI: I didn't hear any questions about the preclinical and clinical. If there are questions from the audience about animal models, et cetera, please bring them up at this time.

DR. WHITE: There are more JRA questions here.

DR. PETRI: Yes, but are there any that are pressing?

DR. RIDER: There is one.

DR. PETRI: Do you want to bring it up now while he is getting ready?

DR. RIDER: Okay. Our question that is fairly

jam

pressing is that most of the emphasis today is on polyarticular course, and yet, there are certain situations where pauciarticular course or even systemic patients would need further study because the agents will be used in those populations.

Also, in order to give a label for all JRA, what sort of representation do we need from these other subsets of patients. So our questions is really how much representation do we need. We will probably not achieve statistical significance out of each subset represented, but how much representation do we need from each subset.

DR. PETRI: Dr. Lovell, do you want to start?

DR. LOVELL: Yes. I think it gets back to the onset and course issue again. I think the label and indication should reflect the type of patients that were studied with that drug and shown to be efficacious, and that may turn out to be pauciarticular JRA patients with uveitis, for example, but that is what the label should indicate.

In all the second-line study, the most reasonable thing would be to characterize the patients as to how polyarticular JRA. So those are the patients that should get the label.

So I think we are probably not going to be able to perform studies large enough to have power in each of the

jam

three subsets, onset subsets, but I think the label really ought to be informed to the clinician by reflecting the patients who are actually enrolled in the study and the entry criteria should reflect kind of the hypothesis that want us to be tested, and then we just make the label fit that rather than kind of arbitrarily constrain ourselves to trying to enroll patients representative of different subsets and that sort of thing.

DR. PETRI: I think this is one of Dr. White's points. What the pediatric rheumatologist care about is the course, not the onset.

Dr. White, did you have anything else?

DR. WHITE: No. I mean, I agree. Absolutely.

DR. PETRI: Does that address your question, Dr. Rider?

DR. RIDER: No.

DR. SCHWIETERMAN: I just want some clarification. Ought the agency provide any guidance with regard to the types of subsets or simply state that the label will reflect the type of data that is derived from the trial design for the JRA label itself? Is that something that is not important?

DR. [FREDERICK] MILLER: Well, the first question that is implied there is that a label for JRA itself cannot

jam

be made. That is your suggestion.

DR. SCHWIETERMAN: Right.

DR. PETRI: We have been told that there are different course. For example, for methotrexate, the pediatric rheumatologist felt very uncomfortable lumping it all as JRA.

DR. LOVELL: I think from a statistician point of view, if I were wanting to get a label for JRA, all types of JRA, for example, methotrexate, it would be difficult to do that statistically.

I think what we ought to try to stimulate in terms of the sponsors is to make sure that they focus their studies with their drugs on the patients that are more reasonable to put on those drugs based on what we know about adults and what we know about JRA, rather than kind of be artificially driven by these three subsets, which I can tell you there is an international criteria that is going to come out for arthritis in children, and there is going to be seven or eight subsets. That seems to be more problematic.

So I think we ought to try to gear the study to the patients at most need.

DR. SCHWIETERMAN: So, if I interpret this correctly, you would be uncomfortable with a generalized JRA claim that didn't describe better in the indications section

jam

the types of patients that were studied.

DR. WHITE: Absolutely, that is correct.

DR. PETRI: So there is a consensus among the pediatric rheumatologist.

Dr. Chambers?

### **Equivalency Trials**

[Slide.]

DR. CHAMBERS: I am going to talk a little bit about equivalence trials and what is equivalence, just how close is close is close enough.

[Slide.]

DR. CHAMBERS: The proposal that is in the document is based on a notion that has been used in the Division of Anti-Infective Drug Products and in some Dermalogics and Ophthalmologics, and it was used historically, originally in anti-infectives when the derm and the ophthalmology products were all in the same division, and so it got carried on.

[Slide.]

DR. CHAMBERS: It has a certain number of assumptions, and those assumptions are when you were doing a comparison that is of a test agent or a particular procedure versus an active control, it also assumes there is no negative control in the study.

jam

If you have a negative control, such as a placebo, or a sham procedure if you are going through a procedure, you have a base mark, some kind of bench mark to go and compare that to, but if you don't have one, then the question was how close did you have to be.

[Slide.]

DR. CHAMBERS: The assumption of the model I am going to talk about right now also assumes that you can determine a success or failure rate for each subject, similar to what we have been doing with ACR criteria composite scores.

It also assumes that all the statistical tests are two-sided, and that the dropouts are handled either as a worst-case scenario or they are treated equally between each of the different groups.

[Slide.]

DR. CHAMBERS: What has been determined what has been used in the past was you drew a 95-percent confidence interval between the test and the control, and if the control agent was a very high percentage, like 92 percent or 93 percent success rate, you said you would be willing to accept your test agent as long as its 95-percent confidence interval stayed within 10 percent of that original. I think it is probably best shown by an example.

jam

[Slide.]

DR. CHAMBERS: If you have got an active control that has a 93-percent success rate, the test agent could be as bad as 83 percent or the 95-percent interval could be as bad as 83 percent, and you would be willing to say yes, my active was 93 and my test agent could be as bad as 83, but it is still close enough to be equal.

[Slide.]

DR. CHAMBERS: That is fine for things that are up high in the nineties. There was some concern that was too strict a criteria if you were not in the nineties. So, if you were between the eighties and nineties, we were willing to look at just how far away you were from 100 and draw the confidence interval around that.

[Slide.]

DR. CHAMBERS: So that, if you had something where the active control was 85-percent successful, you would say you are 15 percent away from 100. So the confidence interval needed to be somewhere between, in this case, 100 and 70 percent. So the lowest you could be would be 70 percent and still be considered equivalent.

[Slide.]

DR. CHAMBERS: Again, there was a question about whether this is too tight once you get to something that is

jam

like 40 or 50 percent successful, but there was also concern -- I'm sorry. This is another example. If you had 82 percent, you could then be down as low as 64 percent. Again, this is 82 percent. It is 18 away from 100. So you could be as far as 82 minus 18.

[Slide.]

DR. CHAMBERS: If you were below 80 percent, at 80 percent or below, then we set this upper limit. This maximum difference you could be was 20 percent because we believe that most people felt that 20 percent away from what the true active control ratio was, was just too far.

[Slide.]

DR. CHAMBERS: That means if you had an active control that was 61-percent successful, your test agent could be as poor as 41 percent and still be considered equivalent.

This is what has been used in the area of anti-infective drug products. It has been used in dermatologic. It has been used in ophthalmologic. It was drawn out of thin air. There is no scientific basis for it, although it has now been used for a number of years.

It made intuitive sense to people because that is what they were willing to say, well, that is close enough based on what I was starting from, and the question we are

jam

asking here is for the types of diseases that we have been discussing within rheumatoid arthritis, is this close enough, is this a valid approach, or does this allow either too much leeway or not enough leeway.

Thank you.

DR. PETRI: If we could open this up for discussion and start with our statisticians.

DR. WOODCOCK: I think, Wiley, you presented this really well, but let me reiterate. This is the confidence limit on the estimate. If there is a lot of variability, the point estimate, the mean or the number of successes has to be very close or even above the control arm.

If using this rule, there was very small variability, which is unlikely in our diseases, then the point estimate could actually be below if you had a very confidence limit. Is everybody following that?

DR. CHAMBERS: That lower bound was for confidence interval, not from --

DR. WOODCOCK: For actual -- what you actually achieved in the trial is not at issue. It is what is the confidence limit around that, and that will vary depending on how much variability there has been on the statistics of the matter, which the statisticians can explain.

DR. PETRI: Dr. Tilley?

jam

DR. TILLEY: There is a lot of methodology that has been developed since those rules came out, and generally, the statistic methodology now asks you to define what you mean by a clinically, in a sense, meaningless difference; that is, how close to two groups have to be to consider that they are equivalent, and then you develop your sample size estimates for your study based on confidence limits around that difference and high power because you don't want to fail to detect a true difference.

So the methodology is different and the way of approaching the confidence limits are different than what has been used in the past. I would recommend particularly in this disease that people look toward that newer methodology.

DR. PETRI: Dr. Chambers, would there be any objection to simply abandoning those old rules?

DR. CHAMBERS: This is up as a proposal. This is not set alone there. I am not sure that the new -- I mean, you still have to make a call of what is close enough, and that is what in these other communities they were unable -- either unwilling or unable to say just how close was close enough.

DR. TILLEY: But I think that is a trial-by-trial decision, just as you decide when you design a trial for

jam

efficacy how much difference you want to detect between your two treatment groups. You have to make some kind of judgment for efficacy. It is the same thing. You have to turn around, and you have to argue something that the clinicians will buy.

Depending on the trial, depending on the agent, there are going to be different values for what isn't clinically important, especially when you weigh the risks in there. I mean, that is part of what people consider when they are designing trials.

DR. CHAMBERS: I guess I would argue that it is not drug-specific; that it is disease-specific. It is how close you want a clinical course to come, not what agent happens to be doing it. So, for rheumatoid arthritis, I think it should be possible to make that call.

DR. WOODCOCK: Let me just say, what Wiley presented was a series of examples. They were really put in the document as examples, so that everyone clearly understands the problem.

Here is the problem, and we do have to draw a regulatory threshold, okay, and I think we agree that it should be based on what are you not willing to lose when you declare that effective -- or compared to another agent, what is a clinically meaningful difference.

jam

DR. TILLEY: But see, it depends on what the other agent is. If you are talking about an agent -- and that percentage does vary based on the success rate of the agent and whether you are talking about a relative or an absolute difference.

DR. PETRI: Dr. Felson?

DR. FELSON: Actually, Dr. Miller also needs to be recognized. He has been having his hand up for a while.

DR. PETRI: Hiding in the corner, Dr. Miller?

DR. [CLINTON] MILLER: First of all, I didn't think that was an arbitrary decision. Maybe I am wrong, but I thought that was developed here at the FDA some years ago in conjunction with some consultants that were basing that rule or that set of rules or variations of that on fiducial limits. Is that not true?

DR. CHAMBERS: I was there at the time, and yes, it was done a number of years ago, but I can tell you, we picked it out of thin air.

DR. [CLINTON] MILLER: It was not a fiducial limit type of statement that you were building on.

Well, the second thing is I think one of the problems with what we are doing now in terms of power is that we are giving a point estimate of power, and I really think what we should see presented to the committee or to

jam

the FDA is a response surface, a power surface where we let differences be there, we allow different sample sizes, we allow different variances, and we build that surface that allows us to develop different criteria for decision-making and see what if this is the case, here is our probability of finding a difference if it is there, et cetera.

What I would recommend as a first step at least would be to look at those power surfaces. People do talk about power curves, but why don't they go ahead and do the job and recognize again what I was trying to say this morning. These are not unidimensional problems. They are multivariable problems, and good decision-making requires that.

DR. JOHNSON: But do you mean that if your analysis showed that it required 500 patients per arm that you should toss it out because the result was too infeasible? Is that what you are saying?

DR. [CLINTON] MILLER: That is correct.

DR. JOHNSON: Okay. Well, we can do that.

DR. [CLINTON] MILLER: Right.

DR. JOHNSON: That, conceptually, is simple to do, but it is actually going at the problem quite differently.

DR. [CLINTON] MILLER: It is called science.

DR. JOHNSON: No, but the logical conclusion of

jam

that is let's take a poll as to what people think the size of the arms of the trial should be, and we can deduce the small difference that should be considered clinically irrelevant. Is that what you are suggesting?

DR. [CLINTON] MILLER: I don't think it is a poll of the sample size. I think it is a poll of the magnitudes of the differences that you want to detect and that it is a function of the variability of the outcomes measures that you have, not a poll of what the "n" is. The "n" can be calculated.

DR. PETRI: Dr. Felson first, and then Dr. Strand.

DR. FELSON: Let me applaud, for one thing, the FDA for beginning to address this because I think this is really a valuable way.

I guess I would even suggest that this is the way that we ought to think about approving drugs in the context in which placebo-controlled trials are difficult or unethical to do.

We can do this now in rheumatoid arthritis or at least, increasingly, we can think of doing it for a couple of reasons. I think we are at the horizon where this is clearly feasible.

Placebo response rates, using the ACR improvement criteria, are now known from a number of trials. So we can

jam

get variability around expected placebo response rate. So we can have a sense of how much, of what percentage improvement a drug needs to have in order to be clearly much better than any placebo would ever be.

Secondly, we know at least one drug, methotrexate, that has been clearly demonstrated in every trial done at this point that it is superior to placebo and it is the standard of care. We could use that as a comparator.

We even have some sense from these trials of where its improvement rate is going to be, depending on the types of patients studied. We could make reasonable estimations of the rate of improvement expected and the variability around that rate. What we are looking at here, really, is the lower bound of the 95-percent CI in comparison. That is the point estimation of interest.

So I think all of this is feasible, and it is a lot more ethical than having people off DMARDs randomize to placebo. So I think it is all doable. I think the data are even beginning to be varying large amounts to be able to make estimates of all these things, and I think based on some of the comments we are having informally that our initial attempts that it ought to be conservative, meaning that we ought to sort of go around this room and say let's be honest here, there is not going to be a placebo in this

jam

trial. We need to have a relatively high bar, initially, so that we are all comfortable that we are in an experimental trial, in experimental treatment that works, without question, even though there is no placebo in this trial.

So, initially, the trials are probably going to have to be either of a terrific agent or very large ones or perhaps both in order for us to begin to feel comfortable with equivalence being used.

DR. PETRI: Dr. Strand?

DR. STRAND: I have several questions. The first is, Dr. Chambers, your proposal is in the absence of a placebo. Is that correct? Because if you do have a placebo arm and you are looking at two active agents versus placebo, then you define equivalency by the confidence intervals which also are determined in part by the difference from placebo for both of the active -- or at least one of the active agents, right?

DR. CHAMBERS: Assumptions that we are making for those were all that you did not have a placebo there.

DR. STRAND: Okay. So the next thing, David, for you is that, of course, we all agree. Active control trials would be wonderful, and methotrexate is our gold standard. Can you tell me what the estimate of the effect size is for methotrexate for a 12-month study by the ACR criteria? Can

jam

you even give me a good way of estimating?

I know that we can say the ball park is, but there isn't a single published methotrexate trial with five of the seven ACR criteria, and certainly not for longer than 24 weeks. So, from that point of view, I would like to say that I applaud this idea, and I think we need to move forward with it, but if we have to prospectively design an active control trial and, at the same time, estimate the effect sizes and what the active control effect is going to be, there is no way to do it that I am aware of.

DR. FELSON: Just to answer the challenge quickly, I think, technically, you are correct in that there is no one that uses all of the ACR core set.

As I said earlier, we have substituted often the group strength for the HAQ, and that is obviously not perfect, but data coming out using HAQ and disability are suggesting very, very comparable rates of improvement.

DR. STRAND: Right, but I don't think the HAQ is the issue.

I am concerned about how you get from 18 or 24 weeks to 52.

DR. FELSON: I think the Mike Weinblatt comparative trial that I pointed out earlier is a 48-week trial, and the CSSRD trials are 24-week trials. I mean, I

jam

don't think we have to mandate a one-year trial here.

I think we can get estimates of the rate of methotrexate response that those are widely varying estimates, and I think we can use those estimates. I think there are other data in the literature that we can get a hold of that also give us more estimates.

DR. STRAND: I understand, but I also think they are too variable, and I do also want to point out that placebo responses in recent studies that have been published using the ACR criteria are, on general, 12 to, say, 18 percent, but there is a couple at 25 and 30 percent, and the CSSRD studies showed us, among other things, that it was the active agent that sometimes determined the degree of the placebo response presumably on the basis, at least in part, on investigator and patient expectations.

So I would like to say that this is a great idea and we are getting there, but I would also like to caution that we are not there yet, and as far as trying to develop a product that could get an approval by some of these suggestions, I think we are all a little bit concerned.

DR. WOODCOCK: I would like to say a couple of things. Wiley, is it true that we could revisit this at another advisory committee meeting?

DR. CHAMBERS: Always.

jam

DR. WOODCOCK: Well, but we had just spoke about that.

DR. CHAMBERS: There is a good possibility we will retalk about some of these things in June.

DR. WOODCOCK: Okay, because, obviously, these issues may require more deliberations, and this was an early introduction.

The agency has in the past looked at active control trials without a placebo arm. These are difficult to interpret. They have used various other kind of rules in the past, which were difficult to interpret.

I don't know that, David -- and this is something we need to discuss -- whether we actually need to impute the placebo effect as long as we are sure that the active control is going to be active, imputing the placebo effect and then calculating what percentage you want to lose as a refinement of this, but we may not be able to do that in the beginning.

DR. FELSON: I agree with you, Janet. What I was suggesting is that the upper bound of the placebo effect that we now know from all of these trials would be a good way to estimate roughly where we want the lower bound of the confidence interval of our new active treatment versus control to be. We want it to be above that level. So none

jam

of us would have any question that this is an efficacious therapy.

DR. PETRI: Dr. Fernandez-Madrid?

DR. FERNANDEZ-MADRID: I would agree with everything that was said, but I think I would caution that if we look at the trials of all the DMARDs that we know from the day one, there have been always some of these trials that have been negative for drugs that we know are effective.

So I think in this type of assay, we need to know that the drugs are active, that this is not really a placebo.

DR. WOODCOCK: The document says one other thing that I want to bring to people's attention, which is that in approval of a totally new agent, we recommend that there be at least one trial that might be a shorter-term trial or one trial where definitive treatment effect is observed, and I think that is what you are getting at.

On this, we seek the committee's advice, that at least one of the trials, one during the development program, there be some demonstration with a difference trial that there is actually a real effect, where these equivalence trials always leave you wondering a little bit.

DR. PETRI: Yes, Dr. Tilley.

jam

DR. TILLEY: I guess I am getting a little confused. Are we trying to say that an equivalence trial is only a test against methotrexate? David was saying we use methotrexate to estimate the effectiveness, but I am not sure that is what you would necessarily want to do.

If you were testing whether your drug is as effective as some other second-line agent that wasn't as effective as methotrexate, which is, of course, Wiley's concern about the creep, I think you would then do your sample size calculations, your estimations a little differently.

So I think we need to be careful when we talk about equivalency that we aren't -- unless we are saying that everyone has to do their trial against methotrexate.

DR. PETRI: I think we are bringing up a practical consideration. Since methotrexate is really sort of the gold standard, clinically, industry is going to have to develop a drug that is equivalent to methotrexate before there will be a market.

DR. TILLEY: Then, that, we should be clear about.

DR. JOHNSON: Well, I mean, or a very safe drug that works half as well as methotrexate. You test it against placebo.

DR. WOODCOCK: If you all are going to be looking

jam

at equivalence trial in approval practices, you want the active control to be something that you believe works reliably in trials if you don't have a placebo arm. So that is not saying it is limited to methotrexate, but something if it is tested repeatedly is going to repeatedly show treatment effect in trials.

DR. TILLEY: And we would need a good, sound rationale for the difference being not clinically important, the difference that the trial is designed to study, and that would have to be provided by the people doing the study.

DR. PETRI: So I believe the committee's consensus is that this methodology does need to be perfected because equivalence trials appear to be very important in RA testing in the future, but we don't have enough information today to make any recommendations.

Yes, Dr. Miller.

DR. [CLINTON] MILLER: I would like to make one observation, if I could. I have an overhead, and it has to do with the claims resulting from these so-called arm experimental designs.

What I want to do is quickly remind our committee and maybe some of the sponsors where there is not a very clear understanding of what interaction means.

[Slide.]

jam

DR. [CLINTON] MILLER: We happened to see a presentation yesterday where, throughout the document, we had a claim of interaction, and it just simply -- you can't prove that from these.

I just want to remind us that the factorial designs and partial factorial designs were created to detect interaction, and if I look, I will call "n" a no-name drug and "m" as methotrexate. I look over here, and these are all imaginary things, but suppose the factorial design is in the upper left corner. We go to the right, and we see some response and we will restrain ourselves to either a log or a linear part of that response. Say that response out there, the unit of "m," is 2. Then, over here in the "n" dimension, I have got a response of 3. So I look at the two-dimensional design space. I have three over there coming from one thing, two in another, and if I look at those squares in the bottom, they are the same squares that I had in the upper left-hand corner.

Now, the question is what happens if the upper left-hand corner in that design space is zero-zero. It is zero "m" and zero "n," and then what happens at "mm"? That is the issue.

Now, if it turns out that that is five, the sum of those two, two and three, that is an additive model. If it

jam

turns out that it is maybe one, well, then I have got interference going on there. If it turns out that it is, like, 10, I've got synergism.

Now, the only way I can measure interaction is to have that true placebo sitting in there. When I do those arm experiments, I have three of those four. So I can't talk about interaction. I have got to come up with the new name for that because it is not -- I am not saying that you can't do that experiment, because you can, and you can compare the "m" and the "n" with an "mm" group, but you don't call that interaction. It is something else, and I don't know that we even have a word for that, except to say that they are significantly different or that they are similar.

That is all.

DR. PETRI: Now I would like to ask the committee's indulgence and push on without a break because I have been told that many of our members are going to have to leave for the airport.

So the next section is going to be on safety analysis and Phase 4 studies, Drs. Miller and Schwieterman.

#### **Safety Analysis and Phase IV**

[Slide.]

DR. [FREDERICK] MILLER: Last but not least, we

jam

are going to switch topics here, and Bill and I will discuss some safety issues.

Before doing that, I just wanted to talk about a few issues that I think override all of the specific safety concerns that we are talking about today. One is that RA is a relatively common rheumatic disorder. Two is that it is a chronic disease requiring long-term therapy. Three, subjects are often taking multiple medications with potential interactions, and of course, most RA patients are women.

[Slide.]

DR. [FREDERICK] MILLER: Although many of these issues actually in the document relate to safety at all phases of drug development, I am going to be focussing on Phases 1 and 2 issues, particularly, and Bill will be focussing on Phases 3 and 4 issues.

I will be talking a bit about trial design and dose escalation, synergized safety, adverse event assessments, stopping rules for individuals and for trials, and Bill will go on and talk about trial size, adequate numbers, trial duration, and follow up possible use of registries, and some special considerations for biologics devices in JRA.

[Slide.]

jam

DR. [FREDERICK] MILLER: Phase 1 safety issues I will particularly discussing here are found in Section 2C, pages 12 to 14 of our document.

Of course, an adequate preclinical safety database is the prerequisite for all clinical trials, and what is adequate depends to a great extent upon the particular specifics of the agent in the population that will be explored in.

An appropriate trial design and the choice of subjects to minimize the risk benefit ratio will also depend upon what type of agents and the toxicities one expects and what types of patients one will be placing into that trial.

In general, we recommend avoiding concurrent methotrexate or other immunosuppressive therapy with the first human use of immunosuppressive agents, and this is for two reasons: one, to avoid confounding the adverse event profile of that new agent; and secondly, to minimize the risk that might occur if there was an unexpected synergistic interaction in immunosuppression.

I think the appropriate initial dose should be guided by the no adverse event dose, estimated by preclinical testing, but often, of course, safety factors of several-fold are useful in this regard, especially if one is dealing with a less-than-very-severe population.

jam

I think conservative dose escalations of half log or less tend to be the rule in many Phase 1 studies.

[Slide.]

DR. [FREDERICK] MILLER: Now, something that is not in our document, per se, but which we in the agency and a number of people outside the agency, particularly the OMERAC task force that has been assigned to look at this issue, is the standardized safety assessment system, and I am going to present this for your thoughts today.

This is defined by a number of features, including a predefined terminology and criteria described in assessed adverse events, systems to optimize adverse event detection, and these include specific assessments that are determined by the patient features and the expected effects of the study agent, the timing of one's assessments, often based on peak drug effect and the potential for longer-term adverse events, and appropriate safety stopping rules.

[Slide.]

DR. [FREDERICK] MILLER: In terms of some of the specific descriptors of adverse events, of course, there are many terminologies here that are in use. We are focussing now on MEDDRA as one of our potential terminologies, but others are being developed.

Adverse event outcome in terms of whether or not

jam

treatment is required, whether or not sequelae occurred, whether they were unresolved or whether death occurred, I think, are important variables to capture, as well as the adverse event severity. These are typically defined as grade 1, mild, grade 2, moderate, grade 3, severe, grade 4, life-threatening.

[Slide.]

DR. [FREDERICK] MILLER: There are a number of these examples of prespecified adverse event rating scales. This particular one comes from NIAID and was developed particularly for AIDS trials, and again, the use of these, I think, actually increases the consistency in adverse event reporting, and I think if any standards could be applied here, it would also help us and other physicians in assessing the toxicities of different agents in trying to help us in our risk benefit ratio considerations.

This particular one, one can use both laboratory values as well as signs and symptoms to try to come up with these different levels of adverse events, grade 1 through grade 4.

[Slide.]

DR. [FREDERICK] MILLER: This often needs to be specifically modified for the particular disease in question, and I think developing one of these for the

jam

rheumatic diseases will be a step forward.

I think that capturing the strength of adverse event association with the study agent is also important. These are often referred to as not remotely, possibly, or probably related, and of course, this is often very difficult to do.

I think often the primary treating physician may be in the best position, the most closely to be able to make these assessments, and certainly, considerations that should be taken into account here include whether or not there was a dechallenge, that is, after withdrawal of the agent, the patient actually improved, or whether there is any rechallenge data, whether there is biological plausibility and whether there is prior reports of this adverse event with this agent or agents like the agent being studied.

The effect of adverse events on the experimental trial, the agent in the trial, was also useful to capture, and this can be, of course -- you may have to alter the dose of the agent or change the dose or you may have to withdraw the patient from the study.

[Slide.]

DR. [FREDERICK] MILLER: Going back to the Phase 1 safety issues in the document, again, it is often useful, I

jam

think, to have adverse event stopping rules, both for individual patients and for the clinical trial, and that is to predict before the trial even begins as many contingencies -- at least the likely contingencies that one might expect in the trial, so that one is not forced in mid trial to try to come up with some decision here.

Of course, the particular stopping rule should be determined by the particular risk benefit ratio for the particular agent and the particular target population. One can accept a higher risk in more severely ill patients.

Often, a grade 3 to 4 adverse event is often used in number of Phase 1 trials for a stopping rule for an individual patient when very little is known about the agent or the duration of its effects.

In terms of the stopping rules for the clinical trial, I think the same caveats apply. If, in fact, one is treating very severe patients, you can accept perhaps a higher risk in that population, but again, often the grade 3 to 4 adverse events in about 5 to 10 percent of the exposed cohort is often used as trial stopping rules, and it is particularly useful when one is dealing with dose escalation studies, and sometimes that is necessary to actually at a certain dose drop back and treat more patients at a previous dose, and sometimes these stopping rules or adjustment

jam

rules, as they are sometimes called, are useful in this regard.

[Slide.]

DR. [FREDERICK] MILLER: Many of the issues in Phase 2 trials are similar to those in Phase 1, but in Section 2D on pages 15 to 17, we talk about some of the Phase 2 safety issues, particularly.

If an adverse event rating scale has not been developed in Phase 1, it is often useful to do that at this stage or to refine it if new adverse events are discovered.

It is useful in Phase 2, of course, to refine the range of safe dosing. It is useful to begin at this point to assess concomitant use with methotrexate or other agents that would be commonly used in the target population.

It is useful to begin to assess the possible risk factors for adverse events at this point as well and to develop a cohort of patients with longer-term follow-up.

It is also important to remember that the trial size impacts not only the confidence in efficacy, but also in safety, and I am indebted to Tony Lachenbruch here for the next two slides, which give one some feeling for estimates and confidence limits, given particular adverse event rates in different size trials.

[Slide.]

jam

DR. [FREDERICK] MILLER: Here is a trial of 100 patients, and you can see that your upper 95-percent confidence limit ranges from about .036 up to .292, depending upon the adverse event rates in that trial. Doubling the size of that trial, essentially, halves these rates, and these are issues, I think, that should be taken into account, depending upon the particular agent and target population that one is going to be looking at here.

So I will close with that and ask if there are any comments or questions about this. I don't have a specific question, per se, to ask the committee regarding this part of the presentation. Dr. Schwieterman does.

DR. PETRI: Dr. Liang had a question.

DR. LIANG: Well, I think if we are going to be sort of thinking about the future, all of us are worried about the mutagenic and other long-term side effects, the immunologic side effects perhaps of the new biologic agents.

We have had patients who have been involved with Phase 1 trials and the T-cells are still down. I think as a sort of physician and potential consumer of these things, the companies that advocate these things have to follow these patients indefinitely, the ones who are getting these major biological agents. I would like to see us actually make a stand on that.

jam

DR. SCHWIETERMAN: Actually, it may be best that I give my short presentation since we are discussing long-term things.

DR. PETRI: Why don't you go ahead.

[Slide.]

DR. SCHWIETERMAN: I didn't mean to preclude the discussion or close it on Fred's items. I think there is a lot to discuss in Phase 1 and Phase 2.

I am going to be presenting the Phase 3 and the Phase 4 aspects of this, but perhaps after this, we could have a general discussion about safety overall.

I would like to just make a few introductory remarks. Because the agency and, in particular, the Center for Biologics is likely to receive submissions in the not-too-distant future where products have higher efficacy rates that have been seen in the past, the overall size of the database that we are going to be receiving may be a bit smaller than has normally been seen. That is number one.

Number two, I think -- and you have already alluded to this, Dr. Liang. The onset of combination therapy, particularly combination therapy whereby immunosuppressive regimens are given and with potentially long-lasting effects, I think is increasingly an issue that we are concerned about.

jam

Because of these considerations and because of the development of the guidelines, we would like the committee's input on what ought to be reasonable guidelines for minimal safety databases with new investigational therapies.

I have divided my discussion into three separate questions, and they cover these particular areas, which will be asked at the end. What are adequate numbers that you need to have to substantiate a safety database? Number two goes right to the question, to the issue that was just discussed, what is the value of Phase 4 registry data, and ought the agency be recommending this more routinely. Thirdly, if there are any comments on the special considerations for biologics devices in JRA, which I will have a few comments on.

[Slide.]

DR. SCHWIETERMAN: Let me just remind you all what the recent ICH guidelines state, namely that 300 to 600 patients treated with a maximum recommended dose for six months and 100 patients treated at this dose for 12 months ought to be available as part of the safety database prior to approval of an agent, and they also make a statement about the total number of patients being treated to be about 1,500. These are for diseases where chronic therapy is needed because they are chronic diseases.

jam

[Slide.]

DR. SCHWIETERMAN: The old 1988 RA guidelines, just for reference, showed something somewhat different that that, that 200 to 400 patients treated for one year, 100 to 200 patients treated for two years ought to be the minimum number for NSAIDs, and roughly, those same numbers, though on the higher end for DMARDs, 400 patients treated for one year and 200 patients treated for two years.

[Slide.]

DR. SCHWIETERMAN: There are many considerations obviously regarding the safety database, and I think we need to make that plain. Overall risk benefit is not simply a numbers game, but also what types of patients you put the drugs into, what the relevant associations are with other classes of agents and so forth.

I have listed a few here that I think are relevant. First of all, just one simple fact that perhaps has already been alluded to, studies with less than 300 patients per group are not power to detect adverse events less than 1 percent, as a general frame of reference, and that is in the document.

The ICH document itself is very clear that there may need to be larger safety databases in the case of certain considerations, and I have listed some of those

jam

here; namely, that there may be late onset ADEs concern because of information from related products, pharmacokinetic and pharmacodynamic properties known to be associated with adverse events, low frequency adverse events, obviously problems with risk benefit, low benefit, of course, being especially problematic if there is a suspicion for toxicity and patients with problems of high background of morbidity and mortality if those patients are included in the study.

[Slide.]

DR. SCHWIETERMAN: I want to say just one word about Phase 4 safety data collection. We have had a lot of discussions within the agency about the value of registries, particularly given some of the concerns that have been raised about long-term immunosuppressive therapy.

The ICH document also comments on this in some regard saying that registries may be useful. They don't say this, per se, but allude to this fact, that if the database is small, if there is late-developing adverse events, including infection and malignancy, questions regarding risk benefit and low-frequency events, more or less the same as I showed you before.

In other words, there may be a role for Phase 4 safety data being collected should questions arise or should

jam

there be insufficient information.

[Slide.]

DR. SCHWIETERMAN: Finally, before I get to the questions, there are some special considerations for biological therapies, devices and therapies for juvenile rheumatoid arthritis.

The document makes it very clear that biological therapies in some cases may have unusual dose response profiles, including high toxicity, narrow therapeutic windows, et cetera, but I also want to make it clear that the comments made in the open public forum this morning were well received, I think, and need due consideration because not all biological therapies behave this way. In fact, many biological therapies, as was pointed out, are not problematic in this regard. Nevertheless, I think that the committee believes, and this is through firsthand experience, that there are many considerations, many instances, rather, where you want to be extremely careful with how you interpret the safety database and so forth, but that this should probably be done on a case-by-case basis, and we are looking into perhaps modifying the wording of the document to clarify this point.

[Slide.]

DR. SCHWIETERMAN: Finally, obviously there are

jam

considerations for devices and for therapies for JRA where the database is likely to be smaller than those recommended by ICH for practical reasons.

[Slide.]

DR. SCHWIETERMAN: So I formulated three questions to the committee, but of course, feel free to comment in general on any of the discussion about safety that we have presented here.

Number one, in general, for drugs in biologics intended for adults, what size is appropriate for a safety database for new agents intended for the treatment of RA? And if the committee could comment on those agents that perhaps have a very high efficacy rate with a low perceived safety problem, what is adequate and what would be recommended as a minimum size?

Secondly, in general, what is the collection of data for registries is useful, and that means ought we to be pushing for those more than we have in the past or are there other considerations that we might be thinking about.

Finally, if the committee completes comment on what size database is appropriate for trial using devices or studying JRA, that would be greatly helpful.

Thank you.

DR. PETRI: Are there general initial comments or

jam

questions before we address each of the questions?

Why don't we start with the first question which is what should the size of the safety databases be. This is really driven by what frequency of adverse events we would wish to detect, and I think, just to start off the discussion, do we want to detect adverse events at the 1 percent frequency or the 5 percent?

Let me ask for committee thoughts.

DR. TILLEY: We are talking about Phase 4 now? We skipped the preclinical? I am not sure where we are.

DR. PETRI: I think we are -- Janet?

DR. WOODCOCK: We are talking about the safety database. We are talking about all the patients, all the exposures that are in the NDA or licensed application, and that won't necessarily all be from controlled trials. Some of the exposers may be short and they may be from Phase 1 trials. That is why the ICH document talks about certain number who were actually exposed for six months or so.

You can accumulate a large number of patients who were exposed for two weeks if you do a lot of short-term trials. So we need advice not only on the whole number of patients, but on exposure.

DR. SIMON: And we are talking about RA.

DR. PETRI: Yes.

jam

DR. LIANG: Are you saying usable patients? I mean, there is attrition when you try to follow patients over time because of various factors. You are hoping that we can spin this out to 20 years. So it is usually 15 percent dropout or lost to follow-up, no matter how good you are. It costs anywhere from 2- to \$6,000 a year per person in a registry is what the general experience is.

DR. PETRI: We haven't even gone to the registry question.

DR. LIANG: No, but I am saying if you are asking usable data over the life of where this question is relevant, you are talking about an inception cohort that could be very sizeable.

DR. WOODCOCK: I guess the committee needs to give us advice. If you have a small preapproval safety database and they impel you to wish for a larger long-term safety database or characteristics of the agent that it may just by its biological characteristics have delayed toxicity or cumulative toxicity is when you might want more of a registry of Phase 4 post-marketing, but I think the question Michelle is asking right now is what is the size. Can you give us some advice on the premarketing side?

DR. FELSON: Well, I guess I had two comments. One sort of harkens back to what Michelle said earlier in

jam

talking about pediatrics, and that is, it depends.

It depends on whether there is a cyclosporin, for example, whether there is a lot of data in use of other drugs, so that we can make the inferences that are reasonable about what we might expect.

It depends whether there are similar agents with similar toxicities already out there that will allow us to make inferences. It depends on the biological basis of the drug and whether perhaps, like in cyclosporin, we might expect actually a number of problems to arise based on how the drug works.

So I guess the answer is it depends. I am not sure there is a generically useful dictate here.

DR. PETRI: But can't we give some guidance in terms of the frequency rate of adverse events that we would want to know at a minimum?

I think an obvious minimum, we would want to detect an adverse event that occurs at a frequency of 5 percent, right? I mean, that would be a bare minimum.

DR. LIANG: Again, it depends on what the base rate is, expected rate, is right?

DR. SIMON: Exactly.

DR. LIANG: It has to be 5 percent over some denominator, and it depends on what you are talking about.

jam

DR. SIMON: And that is particularly true with malignancies in rheumatoid arthritis, for example.

DR. LACHENBRUCH: I think it would help from my viewpoint if we were to talk about what grade of adverse events we are looking at.

I know in vaccines, often you are talking about a 30- or 40-percent rate of grade 1, redness at the injection site, versus a fatality. Clearly, we are far less tolerant of that.

DR. PETRI: Ms. Malone?

MS. MALONE: Yes. This scares me. Really, if you are the 1 percent that it is happening to, you know, it matters a lot. So it depends on the severity of the disease in the population with the risk benefit ratio that you are willing to take. I think it has to be more tailored to depending on what drug and what population you are dealing with.

DR. PETRI: I think all we can give guidance about is a minimum, and that is why I suggested 5 percent as the minimum.

MS. MALONE: But why 5 percent? Why not 1 percent?

DR. PETRI: I think, obviously, if in early clinical testing there were adverse events of 3 or 4, we

jam

would want to detect those with a 1 percent or perhaps even less than 1 percent.

DR. WOODCOCK: I guess the FDA has had experience in approving drugs where after marketing a rare, but very serious or fatal event has surfaced, when the target population is in the hundreds of thousands, perhaps post-marketing. You find out about that real quick, and I guess that is where we are seeking advice.

We are not interested in uncovering a minor adverse event that occurs in 1 percent, but really, we are saying that if you only study 300 people, you could miss an idiosyncratic problems that occurs in 1 in 500. You could miss a recurring problem that is rare, but can occur. That is what we need some advice on.

I agree, and everyone has said it. It depends on how wonderful this new agent is, how much you would tolerate.

DR. PETRI: Dr. Liang?

DR. LIANG: Well, it depends, but I think the other determinants of this decision is its potential dissemination in the general population. I mean, this is sort of the attributable risk guide in public health.

This is a really hot drug. It is only to be given for a rare condition. Well, that is unfortunate, but we are

jam

not talking about disseminating to every household in America.

Sometimes the least toxic therapies have the widest play because physicians, given their druthers, would give something that is as good as placebo and twice as strong. So I think that those are all sort of public health evaluations.

Again, for RA, you would have to think about what potential number of RA patients would see this drug over a certain time in terms of where you would want to set the threshold, I think.

DR. PETRI: I certainly accept your "it depends" answer. Can we get a little more guidance than that?

Dr. Felson?

DR. FELSON: let me just ask the statisticians here, has anyone ever used an empirical bayesian approach to this problem? Because it strikes me this would be appropriate, that one could have preliminary estimates based on animal studies, biological mechanism of action, and early clinical data that would then allow you to estimate within ranges how many subjects you need to have to detect a certain level.

I mean, this would be the right way to do this, not to define arbitrarily for every drug or every agent.

jam

DR. [CLINTON] MILLER: Certainly, there are bayesian models for this kind of thing, but there is another thing; that I think we need to back up just a minute.

You know, it kind of reminds me of the story of the frustrated husband that sits on the edge of the bed and just keeps on telling how good it is going to be. Well, we are sitting here on the edge talking about benefits and about risk, and we never get to the point of actually putting them together.

So, in our minds, we integrate those two concepts, but in fact, we don't do it. Therefore, we get into the arguments, like you had a while ago, that says I am not worried about the time. Well, if you just stopped for a minute and thought about the risk and if your concept of the model was a true risk benefit, you would have to keep the time in that model.

Well, we had the same thing here. Without understanding the risk and the benefits, you can't talk to me very long about what that sample size is going to be. So it looks to me like we are making a real mistake way back there, and it just keeps showing its ugly head as we go ahead and keep on pushing forward without resolving fundamental issues.

DR. PETRI: Dr. Schwieterman?

jam

DR. SCHWIETERMAN: These comments have been very helpful except that I want to make it clear we are not -- well, somewhat helpful -- we have been discussing --

DR. PETRI: We are not inviting you back.

[Laughter.]

DR. SCHWIETERMAN: The agency is faced not with making assessments of risk benefit which all of us would agree would depend upon the data as regards efficacy and safety and all that, but rather, what are reasonable numbers that we ought to guide sponsors with in the vacuum, in many respects, with investigational therapies. In other words, what sort of safety database at a minimum ought these sponsors be using? The reason it is relevant is because the studies are being driven not for efficacy considerations, but for safety considerations. It is hard to know what to tell them when we are not sure what the safety considerations are.

Nevertheless, we think that there is likely to be some minimum number of patients that we would want them to study to exclude a certain event of a certain frequency, and that is what I am asking for.

DR. PETRI: Let me remind the committee, this was something that industry has communicated in their comments as well. They need us to be more specific in our

jam

recommendations.

Dr. Lovell?

DR. LOVELL: I have a question. Has there not been some sort of standard policy at the FDA or in the drugs in the past about the size of databases and using that approach? How often have you been burned in terms of rare events post-marketing? Based on those kinds of practical experience over a long period of time -- granted, it won't apply very well to biologics, perhaps, but for drugs, is there a reason to fix this approach, if there has been a standard approach?

DR. PETRI: The database was outlined by the ICH which is an internationally agreed upon minimum number of patients.

From a practical point of view, most databases submitted to the Center for Drugs have more patients than that, but perhaps the cumulative exposures may not be much greater than that.

Yes, in recent history, there have been very serious catastrophic type of adverse events that have surfaced in the post-marketing period, and this is where the issue of priors -- it didn't happen. It didn't happen in the premarketing database.

At some level, everyone has to be at peace with

jam

this. This is going to happen, but the question is at what frequency. It hasn't happened very much.

DR. LOVELL: I guess my kind of outsider comment is, if the feeling is that it is not broken, why should we go to great kind of arbitrary lengths to try to fix it?

DR. PETRI: We are talking about new biologics, new drugs with new mechanisms where we don't have a big track record the way we do with NSAIDs.

Even with NSAIDs, we have had two NSAIDs withdrawn. Every rheumatologist at this table wrote prescriptions for those NSAIDs which were then withdrawn. Well, Dr. Liang --

DR. JOHNSON: I think what maybe the Biologics people are asking is, you know, you guys think you've got drugs that actually work. In the past, we were so unclear about this that you had to do 10 or 20 trials. So there was a big safety database that kind of accrued by accident.

I think you are saying that the sponsors are saying we don't have to keep driving our plan on the basis of efficacy trials. Isn't that correct? And tell us how many safety exposures we need?

DR. SCHWIETERMAN: That is exactly right, and I think it is a reasonable question that the sponsors are asking us.

jam

I just want to make another point. It is not uncommon in the Center for Biologics to get many -- after the post-marketing phase to get many safety supplements where you identify a whole host of adverse events that weren't even identified at the beginning because of the trial size, because of the trial conduct or whatever, and we in the Center for Biologics in particular are concerned about approving agents with small safety databases where there might be adverse events on a broad range of the patient population, should these agents work well.

On the other hand, we recognize that they are likely to have a high-risk benefit ratio, simply because they seem to work very well.

The question we have, what standards should we set for them in the beginning so that they are properly guided for this committee?

DR. PETRI: Dr. Felson?

DR. FELSON: I wonder if there is a different mechanism for getting safety data than for efficacy data.

If you think about how we are getting safety data, we are getting it from all their trials, which are efficacy-based or dose ranging-based, and you are sort of saying, look, there is a different goal here now. There is going to be a treatment X, biologic treatment X that just

jam

blows everything out of the water, and I think maybe there is such, actually, and that it doesn't require very large sample size to demonstrate efficacy as defined.

It could be done in one or two trials, each with maybe 40 patients in them. So we got a total number of patients over the short term of 80. Yet, we need safety data on a couple hundred people to be confident that we want to release this safely.

Well, what about giving some kind of provisional acceptance to the company to say, look, you have done your efficacy evaluations, let's give you some kind of provisional approveability pending this committee or our formal meeting, but we want more safety data before it gets released? Let them go ahead and a limited number of sites actually evaluate this, get more data on it.

DR. PETRI: David, I feel very uncomfortable with that. I mean, for the most part, rheumatoid arthritis is not a fatal disease.

I think we could have the safety data be before marketing. Let me ask Dr. Luthra.

DR. LUTHRA: I was just going to say I think if you are only talking about biologicals, it may depend upon the drug that is in front of us, and if it is a monoclonal antibody, for example, which knocks out T-cells, we have

jam

already some evidence that this can cause severe infections. It can cause death. Because of that, it can cause malignancies. It can cause other diseases.

Well, it suddenly puts us into a very difficult situation as to what is a risk benefit ratio. There, I think we almost have to have a registry of every patient that is being put on that drug once it comes on the market because you are going to have to follow these patients very carefully.

On the other hand, if you have some sort of a DNA-based protein which is being used like a small peptide and we know with the experience with insulin, erythropoietin, that these are relatively harmless drugs, then suddenly our whole feel of how uncomfortable we are will be able to change.

So I am not sure we can give you an answer. I think it will depend upon what type of a product is in front of us and what are the side effects that have been observed during the studies, and I think here is a situation where you may want to extend the studies for observation for at least a couple of years and not just a six-month study and leave it at that.

DR. PETRI: Dr. Felson first, then Dr. Chambers.

DR. FELSON: Let me go back to your comment,

jam

Michelle, because you prompted me to think about why one might want to -- forget the term "provisional acceptance." Let's talk about availability for uncontrolled use in patients who would like it.

Now, it's very likely that such a treatment being so effective as to need a small number of patients in a trial will be highly desirable for patients who are not especially doing well in their RA. The company and various investigators and patients would be very desirous of getting a hold of it and using it. I think that is probably the case in a number of these.

It would be easy enough to let it be used in that regard without formal approval and use all the data collected from those patients to inform us about safety.

DR. PETRI: But we still don't have a number.

Dr. Chambers?

DR. CHAMBERS: I am going to take us back to the problem, I guess, that I have on, if not daily, every other day. I will have a sponsor come in who has done a Phase 1 trial that has 20 or 30 patients. They are now planning their Phase 2/Phase 3 trials, the whole rest of their development plan. They have not found any adverse events in the first 20 people they did. That is why they are proceeding. They want to know how many patients they need

jam

to go and study in the next trials because they have to go and recruit, get the number of different centers, be able to arrange all of that, something that is going to take six months, a year, two years to go and plan. So they are doing this all up front, and they don't have all of this basic information.

It would be very easy for me to make the determination of how many patients I needed if I had all of this information, all the caveats that everybody is talking about. Not a problem. That is not the situation I deal with every other day.

As has been discussed, there is an ICH guideline, and I think we are looking at that. That is probably going to be a minimum.

There will a whole bunch of caveats along there, and most of the products we are talking about would fit into those caveats. Do we want to go above what that ICH guidance is, and if so, by how much?

DR. PETRI: I think that is where we can say it depends, keeping the ICH guidelines as our minimum. "It depends" would be is there any information about similar drugs. If this is a brand-new biologic that has no past information, then you would want to go above the minimum.

DR. CHAMBERS: I get new classes on an

jam

every-other-week basis.

DR. WOODCOCK: To posit that the ICH could be a minimum, it is going to be required for international registration in many cases. So that may be very reasonable.

What David Felson was talking about, it is perfectly possible for highly efficacious drugs to do open trials or to do safety trials that don't have to have all of the randomization efficacy parameters in everything.

We have tried in our treatment IND programs, which I think most of you are familiar with, where before approval of the drug, it can be released for treatment. To get safety information, we have had a great deal of trouble getting really good safety information because of the way the drugs have been given out, but there is some information that can be gleaned from that. Companies could choose to run safety trials where they are just accumulating patient exposure without having a hypothesis if they think they are going to get their efficacy other places.

If, in fact, you are worried about a toxicity that is also a consequence of the disease, such as disease worsening, you have to do a randomized trial to detect that, and you have to power the trial for safety concerns. So it really depends on the situation.

DR. PETRI: Dr. Fernandez-Madrid?

jam

DR. FERNANDEZ-MADRID: I think I would agree with Dr. Luthra on this subject. I think I would be very conservative. I would require a safety data prior to approval, and I would not limit the detection of the adverse effect to 5 percent.

DR. PETRI: From the audience?

DR. SEAMON: Ken Seamon, Immunex Corporation.

I think one point that is important to consider with respect to what Janet was saying was if one does have a trial which indicates that you have a very efficacious drug, a significantly efficacious drug, with no significant adverse effects, and you have a reasonable population, if you then try and set up a safety trial or some type of limited trial to develop a safety database, it will be very difficult from a sponsor's perspective to keep that trial going given the fact that there is going to be so many people wishing to have availability to that drug.

So, keeping control on the trial, given a very solid database for efficacy, makes it very difficult from the sponsor's perspective.

DR. PETRI: Dr. Abramson?

DR. ABRAMSON: I think I would just like to pick up on what Dan had mentioned. It seems to me that the question of the Phase 2 and 3 is not really where most of us

jam

are uncomfortable with these new biologics or even cyclosporin. It is the late events that may be uncommon.

I guess the challenge, really, is not so much in how to change the ICH guidelines, except maybe on a case-by-case if there was some exception. I believe that pretty much alone would solve the problem of how do you create a surveillance mechanism and who is going to pay for that over a several-year period.

DR. PETRI: All right. I think we are sequeing into the next question, which is the Phase 4 registries, but just to summarize, the consensus was that the ICH should be the minimum and that new drugs with new mechanisms of action might require more than that minimal safety database.

But to go on to this issue of the Phase 4 registries, Dr. Luthra, did you want to just repeat your comment?

DR. LUTHRA: Well, what I said before was that it depends really on the type of biologic agent that we are addressing. If it is a monoclonal antibody and there is enough concern related to monoclonal antibodies right now with several patients who have passed away because of side effects, some whose T-cells are so low they haven't come back in over a year, others who have had overwhelming infections which has led to death, these are all major

jam

concerns, and I think if such a product is to be released, almost every patient should be on a registry.

On the other hand, if we have a peptide, which as we are using for immunotherapy, and we have enough experience with DNA-based -- you know, recombinant DNA-based products like insulin, there I think the safety margins would be very different, and we may not need to have such a tight registry.

Now, I was trying to think beyond this. What would be a way of trying to have some handle? The thought goes through my mind that maybe we should ask the sponsor that the first 1,000 patients after approval should be picked up and followed on a long-term basis. That could be one way of getting a handle because there you could get a risk of .1 percent of a side effect.

DR. PETRI: The only problem I can see with that are the patients who might sort of drop out, no longer taking the agents. You are talking about the first 1,000 patients exposed?

DR. LUTHRA: Right.

DR. PETRI: Or the first 1,000 patients who have had six months to a year on the drug?

DR. LUTHRA: I am thinking about those patients who have been exposed to the drug.

jam

DR. PETRI: But for how long? In other words, I don't think we want the people who just took it for one month.

DR. LUTHRA: Well, but see, some of these antibodies, one shot knocks out the T-cells. They can't get them back. So I think any exposure to that agent, the first thousand patients, we should have them in a registry and follow them whether they stay on the drug or not.

DR. PETRI: Is there any other discussion?

Now, I actually think this is important enough that we should vote on this. So Dr. Luthra's motion was that the first 1,000 patients should be in a mandatory registry. Can I see a show of hands of those who agree?

DR. LOVELL: I think he was talking about depleting antibodies, right?

DR. PETRI: No. We are talking about new mechanism biologics. Remember, he left out the ones based on DNA technology that are thought not to have long-term side effects. This is not going to be all new experimental drugs.

Can I see a show -- Dr. Simon?

DR. SIMON: Just, you know, I have said this before. As a simple country doctor, I am not exactly sure I understand why --

jam

DR. PETRI: Boston is not in the country.

[Laughter.]

DR. SIMON: It is in some country. -- why it is 1,000 as opposed to 1,500. Is there a rationale behind the number?

DR. LUTHRA: Lee, I am trying to look at it. If 1 in 1,000 comes down with something very serious, that is a .1 percent risk of capturing that incidence.

You know, we can always go to the first 100,000 if you really want to go wild about it, but the whole point is we have to be rational as to what the number should be.

Now, I am not sold that 1,000 is the final number. If the group decides 5,000, that is fine with me, but we are trying to be kind of rational about it.

DR. SIMON: I just wondered whether or not the proposal should not be a number, but it should be establishing a registry that would be appropriate for what we are particularly looking for and that it is the registry issue that is the key issue.

DR. PETRI: All right.

DR. SIMON: Not the number.

DR. PETRI: I am actually very willing to accept that.

Dr. Schwieterman, you had a comment?

jam

DR. SCHWIETERMAN: Well, since Dr. Simon's comment, I really don't. If we were going to vote --

DR. PETRI: Those country doctors get you every time.

Dr. Strand, and then Dr. Abramson.

DR. STRAND: Well, I wanted to actually respond to two things. First of all, making biologics particularly available or even new drugs on treatment INDs is wonderful, but as Janet pointed out, you don't get very good safety data because you end up getting your entire clinical research department overwhelmed with telephone calls, shipping drug, and the case report forms not only don't get filled out, but they don't get collected because you are still shipping drug to new patients. It becomes, actually, unfortunately not a very good way to do it.

It is good to take people who are successfully treated and allow them to get continued open-label administration.

In terms of registries, I think we talked about this yesterday. We don't know whether it is the drug effect, the biologic effect, or the underlying disease, and to me, a better proposal might be that you can keep a registry on the patients that you have treated. It may not be a thousand or 1,500. It really depends, but the other

jam

point would be that we could, as rheumatologist through, say, the toxicity working group or something, set up the study that Brian Strom proposed yesterday, which would really get at rheumatoid arthritis patients right now treated with current agents and might help us know what is ultimately happening in terms of lymphoproliferative diseases and some of the other concerns.

If we did that through sort of a working group, then all the companies could support it, and it wouldn't become a prohibitive cost. If you are setting up new registries, it can be very, very expensive, and you don't necessarily have interpretable data. We had 1,300 patients in the azathioprine registry in Canada before we could actually make a more accurate estimate of what was going on with azathioprine in RA.

DR. JOHNSON: Would that registry capture drugs before they are approved, though? I mean, somehow after the fact, once that drug is approved, that whole cohort of patients get tossed into the registry. Is that what you are saying?

DR. STRAND: They certainly could be added to the registry. I mean, it is not really a registry that Brian Strom was presenting. It was more like a huge epidemiologic survey, but it would go after, say, 50,000 or 100,000

jam

patients, and they would determine the diagnosis of RA and then they would be looked at every six to 12 months for the next long period of time, and if we did it on a cooperative effort, it would not be very expensive and it would be available to everyone, and I think we would learn a lot more of what we are trying to get at.

I don't think we honestly know if it is the disease or the products.

DR. PETRI: Dr. Liang?

DR. LIANG: I don't think that the attribution problem, irrespective of whether Brian Strom does it epidemiologically or we do it in a prospective cohort -- and I think what I am looking to this as is a sentinel cohort of people who have early experience that we can follow on the time, and it would be an early warning signal if there is some rate of either death, infection or malignancy.

I think that even when you did that, you would still have to do the typical thing, which is to do an analysis within that cohort, but this would be -- this is a prospective way of doing something that Brian would do on a cross-sectional way.

DR. PETRI: Dr. Abramson?

DR. ABRAMSON: I think this is a very important issue, but I am troubled by the generalization of the

jam

biologics, once again.

I think this has to be a case-by-case situation. If anti-TC-cells deplete T-cells, that is a serious concern, but I have concerns about cyclosporine for prospective studies. I have concerns about new immune modulators that we haven't seen yet. So I like the idea of the registry and all of these issues, but the proposal, as stated, was all biologics, I think.

DR. PETRI: What I tried to do was to move the ones for which there are no safety concerns. So perhaps we should rephrase it that new experimental drugs for which there are safety concerns and that we would ask for Phase 4 registries, but we are not going to mandate the number of patients that have to be in those registries.

DR. TILLEY: Then, that would not preclude those patients being part of this -- you know, the registry being a part of the bigger effort that you were describing. So the registry could be anyplace. It wouldn't necessarily --

DR. PETRI: No. Of course, it doesn't preclude there being cooperative efforts, but since those aren't currently in place --

DR. STRAND: Well, we are trying to set them up, and what I am trying to get beyond is this idea that you automatically have to take the first thousand patients on a

jam

new drug and follow them because, in fact, you don't know that you really are going to get the kind of quality information that you want from those thousand patients.

If we set up a mechanism, we are going to be more likely to get the actual information we want. I mean, it is all very nice to say you are going to do that, but the follow-up and the recall of patients and so on and so forth is pretty hard.

Plus, if you have a thousand patients, you haven't seen a rare adverse event, then you say you don't need a registry for this particular product, and at patient 1,002, there is your rare event.

DR. PETRI: That is always going to be the limitation of this, but at least the registry is a first step, and I think that is the committee's consensus is that we do need to take some sort of first step.

DR. JOHNSON: Is there any regulatory dimension to this, Janet? Can we mandate a five-year report on a registry?

DR. WOODCOCK: These are commitments that are made by sponsors prior to approval of drugs or biologics. There is no way to -- post-marketing, there is no way for the agency to insist that this get done. So we have had some problems.

jam

Actually, there are a lot of logistical problems in doing these, and so there have been some problems in getting folks to actually come through with their commitments, and we have had problems in actually following the adherence to the commitments. So this is not a fool-proof mechanism of getting follow-up.

I would say that, in general, I would think when registries are most useful for events that have a time, their secular trend to their occurrence, their cumulative toxicity or something like that, they require a lot of drug exposure or something. You can handle rarer events just by looking at a lot of patients.

DR. PETRI: Additional comments before we rephrase the question?

Dr. Liang?

DR. LIANG: Well, I mean, I recognize the practical limitations. In fact, we have sort of glibly dismissed a lot of our problems yesterday and today with PMS or registries, and they are nontrivial to do them right and to do them well.

On the other hand, there are some novel ways of getting this data, for instance, using administrative data. Perhaps if there was a law that required that before you got paid through an insurance company that any kind of drug like

jam

this were part of the billing form, you could do some interesting things, but I think it is the concept, not the specific thing of a registry that we should try to explore ways to make it practical and feasible to track patients who have been exposed to these agents for as long as we can.

DR. PETRI: So, to try to rephrase this, we would recommend that there be a mechanism to set up Phase 4 registries for new experimental drugs where there are safety concerns. I am leaving this very open-ended.

DR. LIANG: It is hard to disagree with that. So maybe we should vote on it right now.

DR. SIMON: But only those with safety concerns?

DR. LIANG: I think that all of us have anxiety that we are in the unknown here and that what we think are the major toxicities, we would like to just count noses indefinitely.

DR. PETRI: We backed off from a lot of the specifics. So I am hoping we can reach consensus at this point.

Can I see a show of hands for those who agree with the question as phrased?

[Show of hands.]

DR. PETRI: And dissenters, please raise your hands?

jam

[No response.]

DR. PETRI: So there is no dissent.

The next question is on registries for devices and for JRA.

DR. LIANG: May I speak?

DR. PETRI: Dr. Liang?

DR. LIANG: I think the devices is a tar baby of major proportion, and the reason I am saying this is because we have been particularly interested with hip implants and what has happened over the years.

A nickel's worth of our findings is that basically all the laws are in place to do "post-marketing surveillance." It doesn't work, and we know out there in the hinterland that people have had implants that have fallen apart after several years which are clearly worse than the older models, and there are hundreds and hundreds of implants.

I think that the document doesn't really give the full texture of the problem irrespective of whether we are talking RA or OA, and I think if we are going to do it, we have to do it much better. I think we can't just do it from the perspective of rheumatoid arthritis. It is a generic problem.

DR. DAWISHA: Sahar Dawisha, Center for Devices.

jam

I wanted to address your comment, Matt, and I want to preface what I am going to say by the fact that there are currently no medical devices approved for the treatment of signs and symptoms of RA, and the purpose of this document is for the therapy of the patient and therapy of RA. So products such as hip implants would not really fall under this particular guidance document.

I also wanted to just make a comment about safety databases in devices. In general, it has been a problematic area, but one of the ways that we have approached this is through post-approval studies, which is essentially prior to marketing while the Phase 3 study is ongoing or during the approval process, there is essentially a Phase 4-like database that is collected, recognizing that the entire duration of follow-up wouldn't be reached by the time approval is made, but that at least there would be efforts in that direction.

DR. LIANG: May I ask a question? Why is this not considered treatment for rheumatoid arthritis? It is one of the most effective things we do. I don't understand the word meaning here.

DR. DAWISHA: I guess for several reasons. One is that it doesn't -- when you are putting in one hip implant or one joint implant, you are not necessarily treating the

jam

signs and symptoms of RA in terms of the traditional definition of the ACR core set, for one thing, and then, for another thing, this is intended for products that are going to be coming up for approval. As you know, hip implants are already approved devices.

DR. PETRI: Dr. Simon?

DR. SIMON: Well, actually, this came up when we were dealing on a devices panel related to an injectable material that was under the devices characteristics.

Now, admittedly, that also was not going forward under approval for rheumatoid arthritis, but the discussion had a lot to do in the same fashion that we just had with biologic materials. Clearly, I don't know this for sure, but I'm sure there must be a document in production about OA, similarly like this document, and I would suggest that under those circumstances that long-term follow-up was the major issue about this particular device.

We had no idea what would have happened based on recurrent injections on this particular material or what would happen even after five years having had the injections done. My concern is that the thematic should be what we approve. The idea is a good one, to do a follow-up and to make sure it gets done and assure that the sponsor understands that is their responsibility when coming in with

jam

an NDA or whatever they call the new devices proposal.

So I would like to urge us to consider this as an important thematic that we want to support.

DR. WOODCOCK: I think it would be imprudent for this committee not to vote on the device question. We are interested in your comments, though, on this.

DR. PETRI: I think this is just an extension of our previous vote. I mean, the whole consensus of the committee is that you need to have follow-up data for safety issues.

As we are winding down, Kathleen Reedy has a reminder to us, and then Janet will have some closing comments.

MS. REEDY: As you are packing up and getting ready to leave, thank you for coming, and if you would like to take your materials home, please do. If you would like for us to recycle them for you through the shredder, please put them on this round table here. If you would like anything Fed-Ex'd back to you after you leave, put a note on the materials and leave them in front of you. Thanks.

DR. PETRI: Janet?

### **Conclusions and Summary**

DR. WOODCOCK: Well, I would like to thank the committee. I would like to thank the Chair for bravely

jam

running this meeting.

People have remarked to me how engaged everyone has been on this document. We really have had some extremely substantive input into our deliberations.

I think the input has been so substantive, in fact, that I believe you will be seeing another draft of this. You may not have to have another marathon meeting like this, but I think we will redraft this document and show it to the committee before we would issue it in final because there are a number of open questions that remain.

Thanks to the audience as well for their participation.

DR. PETRI: We are adjourned.

[Whereupon, at 4:14 p.m., the Advisory Arthritis Committee meeting concluded.]

- - -