Guidance for Industry

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

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiologic Health (CDRH)
February 1999

Clin 7

96D-0067 GDL 3

Guidance for Industry

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

Additional copies of this Guidance are available from:

Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane, Rockville, MD 20857
(Phone 301-827-4573)
Internet: http://www.fda.gov/cder/guidance/index.htm.

or

Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
Internet: http://www.fda.gov/cber/guidelines.htm.
Fax: 1-888-CBERFAX or 301-827-3844
Mail: the Voice Information System at 800-835-4709 or 301-827-1800.

or

The Division of Small Manufacturers Assistance (DSMA), HFZ-220
Center for Devices and Radiological Health
Food and Drug Administration
1350 Piccard Drive, Rockville, MD 20850
800-638-2041 or 301-443-6597
Internet: DSMA@CDRH.FDA.GOV
Fax: 1-301-443-8818
Facts-On-Demand (FAX) at 800-899-0381 or 301-827-0111

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
February 1999

TABLE OF CONTENTS

I. INTRODUCTION 1

II. NEW CLAIMS FOR THE TREATMENT OF RA 2
- A. Reduction in the Signs and Symptoms of RA 2
- B. Major Clinical Response 3
- C. Complete Clinical Response 4
- D. Remission 4
- E. Prevention of Disability 4
- F. Prevention of Structural Damage 5

III. CONSIDERATIONS IN RA PRODUCT DEVELOPMENT 6
- A. Preclinical Considerations 7
- B. Pharmacokinetic/Pharmacodynamic Strategies 11
- C. Considerations in Phase 1 Trials 11
- D. Considerations in Phase 2 Trials 14
- E. Efficacy Trial Considerations 16
- F. Safety Analysis 30

IV. SPECIAL CONSIDERATIONS FOR BIOLOGICAL PRODUCTS 32
- A. Species Specificity 32
- B. Dose Responses 32
- C. Toxicity Response 32
- D. Product Homogeneity 32
- E. The Role of Antibodies 32

V. SPECIAL CONSIDERATIONS FOR MEDICAL DEVICES 33
- A. Background 33
- B. Efficacy Considerations 33
- C. Safety Considerations 34

VI. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS 35
- A. Background 35
- B. Applicability of Pediatric Regulation and Impact on Trial Design for JRA Studies 36
- C. Outcome Variables and Claims 37
- D. Trial Design Issues 3 9
- E. Concurrent Antirheumatic Agent Administration 40
- F. Multicenter Trials and Center Effects 40

REFERENCES 42

APPENDIX A: COMPARATIVE TRIAL RESPONSE RATES 45

Guidance for Industry[1]

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

I. INTRODUCTION

This guidance is intended to assist developers of drugs, biological products, and medical devices intended for the treatment of rheumatoid arthritis (RA). The document discusses the types of label claims that can be considered for such products and provides guidance on the clinical development programs to support those claims.

The central purpose of label claims is to inform prescribers and patients about the ***documented*** benefits of a product. Because RA is a chronic, symptomatic disease that can result in a variety of outcomes with different chronologies, severities, and overall patient effects, any number of different clinical outcomes could provide the basis for a label claim.

Relief of symptoms — the *signs and symptoms* claim — is a central therapeutic effect of most RA therapeutics marketed circa 1997. The claim structure proposed in this document, however, incorporates a wider range of patient outcomes than previously allowable RA claims. As a result, guidance is provided for demonstrating patient benefit of greater magnitude than is needed for a claim of symptomatic relief. For example, the claims *major clinical response*, *complete clinical response*, and *remission* (the same criteria as *complete clinical response* while off all antirheumatic drugs) reflect enhanced effects on the signs and symptoms of disease. The claim *prevention of structural damage* is documented by various radiographic techniques. The claim *prevention of disability* is intended to reflect longer term benefits on disease course. The claims and clinical development programs discussed in this draft guidance for industry represent the current views of Agency rheumatologists about achievable and clinically relevant overall outcomes that can be evaluated in clinical trials.

Traditionally, RA therapeutics have been categorized as *disease modifying antirheumatic drugs* (DMARDs) or as *nonsteroidal anti-inflammatory drugs* (NSAIDs). As a result of the ongoing search for more effective therapeutics that have a positive impact on the natural history of the disease, promising new therapies are currently being tested in the clinic. Many of the novel agents

---

[1] This guidance has been prepared by the Rheumatology Working Group of the Medical Policy Coordinating Committee (MPCC) of the Center for Drug Evaluation and Research (CDER), the Center for Biologics Evaluation and Research (CBER), and the Center for Devices and Radiological Health (CDRH). This guidance document represents the Agency's current thinking on Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Rheumatoid Arthritis. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

under study for the treatment of RA defy categorization by putative mechanism of action. As a result, the usefulness of classifying them in the traditional manner may be limited. For this reason, information being provided in labeling about the onset and duration of action and the durability of response of therapeutic interventions is intended to reflect the data that were gathered in clinical trials. Because of this, some of the claims described in this document incorporate response duration times within their structure.

Over the past decade, there has been a search for better measures to describe patient outcomes in RA clinical trials. A number of organizations, including the European League Against Rheumatism, the International League Against Rheumatism, the American College of Rheumatology (ACR), and the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group, have attempted to define core groups of measures, as well as composite indices, that describe patient outcomes. As a result of these efforts, several new measures have been described and validated with clinical data. With the hope that these measures will provide more useful information about patient outcomes, this document provides guidance about the use of these new measures in clinical trials to support label claims.

One outcome measure that is not fully relied upon as a stand-alone claim is general *health-related quality of life* (HR-QOL). Since RA affects so many domains of a patient's life, it is hoped that such HR-QOL measures may provide an integrated assessment of the long-term impact of intervention. However, not enough information is available on the performance of general HR-QOL measures in longer term arthritis trials. The incorporation of such measures in planned trials is encouraged.

II. NEW CLAIMS FOR THE TREATMENT OF RA

A number of new claims are now being evaluated in clinical trials during drug development. Descriptions of the claims and acceptable outcome measures to support each claim are discussed in the following sections.

A. Reduction in the Signs and Symptoms of RA

This claim is intended to reflect the demonstration of symptomatic benefit or benefits that includes improvement in signs of disease activity as well as symptoms. Reduction in signs and symptoms may ordinarily be the initial claim granted for marketing approval. Ordinarily, this claim is established by trials of at least six months' duration, unless the product belongs to an already well-characterized pharmacologic class (e.g., NSAIDs) for which trials of three months' duration are sufficient to establish efficacy for signs and symptoms. Six-month trials are desirable for several reasons. First, RA is a disease of long duration. Interventions that provide only short-term, time-limited benefit are unlikely to have overall value to patients. In addition, products with the potential to elicit antibody formation should be assessed for durability, since antibodies may block effectiveness. In evaluating signs and symptoms, methods that evaluate response over time are preferable to methods that incorporate only the baseline value and the final observation, unless there is a

reason to weight symptoms at the last visit more than intermediary symptoms. Acceptable outcome measures that would support claim A include:

1. Validated composite endpoints or indices of signs and symptoms

These composites may be used to construct categorical endpoints for patient success or failure. For example, the Paulus criteria (1990) or the ACR definition of improvement (ACR 20)[2] could be used to assess a patient's response.

Illustration: Success for each patient in a six-month trial could be defined as meeting the criteria for improvement over baseline in at least four of six monthly observations and not dropping out because of toxicity.

2. Well-accepted sets of signs/symptoms measures

The four measures previously recommended in the 1988 *CDER Guideline for the Clinical Evaluation of Anti-Inflammatory and Antirheumatic Drugs* (FDA 1988) — joint counts: pain, tenderness, and swelling and global assessments: physician and patient — or the ACR core set are examples of well-accepted sets of signs and symptoms measures. The criteria for success and the methods for statistical analysis should be prospectively defined and agreed upon. For example, historically, in using joint counts and global assessments, a statistically significant difference between the control and the treatment group in change from baseline in at least three of the four measures has been used as the criterion for a successful trial. However, as stated above, comparison of only the baseline and last observation may not be the best way to construct the analysis since this method leaves out all intervening efficacy observations.

For both the above measures, using 66- or 28-joint count is appropriate (Smolen 1995).

B. Major Clinical Response

This claim is intended to reflect the demonstration of a continuous six-month period of success by the "ACR 70," which is defined entirely parallel to the ACR 20, except 70 percent improvement, rather than 20 percent, is needed for the component assessed. This claim is based on statistically significant improvement in response rates by the continuous six-month ACR 70 definition compared to background therapy in a randomized control group. For reference, the number of patients satisfying various definitions of ACR

[2] The ACR definition of improvement (ACR 20) is 20 percent improvement in tender and swollen joint counts and 20 percent improvement in three of the five remaining core set measures: patient and physician globals, pain, disability, and an acute phase reactant (Felson 1993, 1995). For specific details on individual measures, e.g., disability, acute phase reactant, refer to Felson 1995.

responses from ACR 20 to ACR 70 in two historic databases are given in Appendix A of this document. Trial duration should be a minimum of seven months for an agent expected to have a rapid onset of action and longer for agents with less prompt effects.

C. Complete Clinical Response

This claim is intended to describe a therapeutic benefit of greater magnitude than the *major clinical response* claim. *Complete clinical response* and *remission* (see below) are identically defined as a continuous six-month period of both *remission by ACR criteria* and radiographic arrest (no radiographic progression [Larsen 1977] or modified Sharp methods [1985]). Complete clinical response connotes a benefit requiring ongoing drug therapy; *remission* is defined by the same result while off all antirheumatic drugs. The 1981 ACR remission criteria (Pinals 1981) require at least five of the following: morning stiffness less than 15 minutes, no fatigue, no joint pain by history, no joint tenderness or pain on motion, no swelling of joints or tendon sheaths, and erythrocyte sedimentation rate (ESR) less than 20 for males or less than 30 for females. The duration of trials designed to support this claim will vary depending upon the rate of onset of effect of the test product. For all but the most rapid-onset agents, trials of one year's duration should be planned. Longer trials may be needed for very slow-acting agents. Trials evaluating complete clinical response would use a categorical endpoint (patient complete response or treatment failure) as the primary outcome measure.

D. Remission

This claim is defined as both *remission by ACR criteria* and *radiographic arrest* (no radiographic progression by Larsen or modified Sharp method) over a continuous six-month period while off all antirheumatic therapy. Remission is not intended to imply cure, and a remission claim could be granted even if patients relapse after six months or more of remission. The duration of trials designed to support a remission claim will vary, depending upon the rate of onset of effect of the test product. Ordinarily such trials should be at least one year in duration, and longer trials may be needed for slow-acting agents.

E. Prevention of Disability

This claim is intended to encourage long-term trials in RA. Currently, the Health Assessment Questionnaire (HAQ) (Fries 1982) and the Arthritis Impact Measure Scales (AIMS) (Meenan 1982) are adequately validated measures for use as the primary outcome measure in these trials. Studies should be two to five years in duration. Sponsors seeking this claim should plan to have demonstrated previously, or to demonstrate concomitantly, improvement in signs and symptoms. Since the full effect of RA on a patient is not captured without the use of more general HR-QOL measures, a validated measure such as the SF-36 should also be collected and patients should not worsen on these measures over the duration of the trial.

F. Prevention of Structural Damage

Prevention of structural damage is an important goal of RA therapy. Trials evaluating this outcome should be at least one year in duration.

The following are examples of outcome measures that could be used to support prevention of structural damage claims.

1. Slowing X-ray progression, using either the Larsen, the modified Sharp, or another validated radiographic index

   Radiographic claims should be based on comparisons of films taken at one year (and subsequent yearly points) with those taken at baseline. All randomized patients should have films at both time points, regardless of whether they are continuing treatment. Patients dropping out of the trial should have films taken at that time. Prespecification of the handling of dropouts is especially important in these trials.

2. Prevention of new X-ray erosions — maintaining an erosion-free state or preventing new erosions

   Trials evaluating this claim would ordinarily use a categorical endpoint to assign a status of progression or nonprogression to each patient, comparing the final state to the baseline state.

3. Other measurement tools (e.g., MRI)

   Other measures, such as MRI (magnetic resonance imaging) or ultrasonography, could be employed. However, because of the technique's potential for identifying small, albeit statistically significant changes, the magnitude of the difference that would reflect actual patient benefit is unclear and needs to be established.

Because slowing of radiographic progression does not in itself define a patient benefit, it is expected that the claim of prevention of structural damage would be submitted for an agent that has been shown (previously or concomitantly) to be effective for one of the other claims (e.g., prevention of disability). However, some agents are not intended to affect acute inflammation, but are designed to prevent or slow joint destruction by other means. The first indication that such an agent is clinically useful might be slowing of radiographic progression. Nevertheless, the ultimate goals of slowing joint destruction are to improve symptoms and to preserve functional ability. Therefore, slowing radiographic progression of disease is considered a surrogate marker for overall patient benefit in RA.

Under 21 CFR 314, subpart H and 21 CFR 601 subpart E, FDA can approve drugs intended to treat serious and life-threatening diseases based on an effect on a surrogate marker, provided that certain criteria are met and that there is a commitment to define the

actual clinical benefit of the agent in studies completed after marketing. A demonstration of significant slowing of radiographic progression in a seriously affected population of RA patients would qualify for consideration under these regulations. Sponsors are urged to consult with the relevant FDA staff before embarking on a clinical program based on these regulations.

One example of a significant effect on radiographic progression might be the demonstration, in a randomized controlled trial, of maintenance of an erosion-free state in a large majority of treated patients when control patients develop multiple erosions. The methods of measurement of the radiographic finding, the magnitude of change considered to be clinically significant (whether a by-patient measure or treatment group means), and the methods of statistical analysis should be prospectively defined in the clinical protocol, and the sponsor should seek Agency concurrence with the plan for evaluating efficacy. The use of the *accelerated approval* pathway would necessitate timely completion of phase 4 studies using acceptable clinical endpoints evaluating signs and symptoms or prevention of disability. It is anticipated that these investigations would be extensions of the one-year studies used for the *accelerated* approval.

III. CONSIDERATIONS IN RA PRODUCT DEVELOPMENT

The following information on preclinical and early clinical product development pertains primarily to pharmaceuticals (drugs and biologics). Except in the first two sections, the general principles outlined below also apply to devices. For information specific to the development of devices, refer to the section in this document entitled "Special Considerations for Medical Devices." Developers of products that combine therapeutic modalities (e.g., biologics and devices) may request assistance from FDA in designating a lead center for review of the product. Such requests should be submitted to: Office of the Chief Mediator and Ombudsman (HF-7), Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857.

Frequently encountered issues in RA product development include:

1. Selecting appropriate in vitro (animal or human systems) and in vivo animal models for screening potentially active agents.

2. Designing and performing appropriate preclinical safety studies to support the use of a new molecular entity in human volunteers or patients.

3. Balancing the potential need for therapeutic intervention early in the disease course with the need to avoid exposing patients with mild disease to agents that have toxicities or little record of safety.

4. Identifying the potential risks associated with combination therapies, particularly those with shared target organ toxicity or potential for pharmacokinetic interactions.

5. Designing adequate and practical long-term safety monitoring.

6. Designing trials that definitively show clinical efficacy.

The following sections discuss approaches to the above issues.

A. Preclinical Considerations

This section focuses on preclinical issues that are specific to the clinical development of antirheumatic therapies. In designing toxicity studies and the timing of such studies, consultation with the Agency is recommended concerning the current recommendations and guidances that address drugs, devices, and biological products. Guidance on preclinical safety testing, addressing the need for and design of toxicokinetic, reproductive toxicity, genotoxicity, and carcinogenicity studies has been developed by the International Conference on the Harmonization (ICH) of Technical Requirements for Pharmaceuticals. Because biologics can pose unique challenges in animal study design (for example, species-specific binding or immunogenicity of human proteins in animals), a specific ICH document is available that addresses the safety evaluation of biotechnology-derived pharmaceuticals (ICH S6 1997).[3]

1. Pharmacokinetics

Animal studies of drug absorption, distribution, metabolism, and excretion are important during the early investigational new drug (IND) phase to aid in toxicity study interpretation, but need not all be completed prior to phase 1. Generally, for initial studies in humans, determining pharmacokinetic (PK) parameters, such as area under the curve (AUC), maximum concentration (C_{max}), and half-life ($t_{1/2}$) in animals, is sufficient to provide a basis for predicting safe clinical exposure.

In the past, preclinical testing of combinations of drugs (or biologics) to be used in patients with RA has not often been done prior to the initial clinical trials. However, given the variety of drugs, including NSAIDs, analgesics, corticosteroids, and DMARDs currently used to treat RA patients, it would be useful to consider this testing of common combinations both preclinically and clinically. In addition, to evaluate potential interactions, information on the impact of concomitant therapies on pharmacokinetics may be needed to optimize dosing regimens and to identify potential safety concerns. Metabolic interactions often may be assessed in an in vitro system using animal or human liver slices, microsomal preparations, or purified P450 enzymes (FDA 1997).

Interactions may also result from the presence of individual- or disease-specific factors, such as rheumatoid factor, which may bind to various

[3] ICH documents are available via the FDA Internet home page at http://www.fda.gov/cder or cber.

monoclonal antibody therapeutics; in such cases in vitro binding studies that identify patients with high titers may be useful in identifying patients who may exhibit unique pharmacokinetics or patterns of clinical response.

2. Biological activity

The biological activity of a potential antirheumatic therapy should be established using multiple preclinical model systems (i.e., in vitro, in vivo, ex vivo). In vitro screens can use cells or tissues derived from animal or human sources and are generally used to select candidate drugs that have a desired effect on a molecular target. Such assays can also be used to devise appropriate bioassays for the selected agent. Animals, either healthy, with rheumatic disease (spontaneous or induced), or genetically modified, are subsequently used to determine whether the biological effect can be demonstrated in vivo. While the in vivo system used should mimic one or more aspects of rheumatoid arthritis or its etiology, it is expected that each animal model will have limitations.

a. In vitro

Data from in vitro studies can be useful in defining the potential mechanism of action of a drug or biologic and for determining relevance of a particular animal species for in vivo assessment of activity or safety. These data are especially useful if a potential surrogate marker can be identified in preclinical studies. For example, if the product is intended to affect the CD4 receptor on lymphocytes, this receptor can be used as a surrogate marker for both activity and certain toxicities.

Several in vitro tests could be used, depending on the mechanism of action of the drug or biologic. For example, binding assays may be useful for developing receptor antagonists or monoclonal antibodies. In vitro functional assays (e.g., platelet and neutrophil aggregation) may be useful tests for identifying inhibitors of inflammatory mediators. Enzymatic assays (e.g., in vitro or ex vivo inhibition of cyclooxygenase, lipoxygenase, and phospholipase) may also be useful for determining selectivity for the inhibition of isozymes.

b. In vivo

Selection of animal models should be made on the basis of pharmacodynamic (PD) responses, similarity of animal disease etiology to clinical disease, and/or to define mechanism-based toxicity. Ideally, products that are targeted for a subset of arthritic patients should be developed in an experimental model(s) that is most relevant to that subset. For example, rats are not sensitive to drugs that inhibit 5-lipoxygenase.

Therefore, mouse or rabbit models are more relevant to evaluate the anti-inflammatory activity of leukotriene inhibitors.

The development of rheumatic disease models to allow screening for potential RA candidate drugs is encouraged. The following examples are meant only to illustrate some models in current use and are not intended to suggest excluding the use of others.

Collagen-induced arthritis (CIA):

Collagen-inducted arthritis is often considered to be a suitable model for studying potential drugs or biologics active in human rheumatoid arthritis because of the involvement of localized major histocompatibility, complete class II-restricted T helper cell activation, and the similarity of histopathological lesions. Radiographs of joints affected by CIA often show erosive changes similar to those seen in human rheumatoid arthritis. The progressive arthritis often results in RA-like joint deformity and dysfunction. Anticollagen antibodies, which occur in some patients with RA, develop in the CIA model.

The CIA model has been useful for assessing immunosuppressants and steroid hormones as well as inhibitors of inflammatory mediators. Since this model can be induced in several animal species, it may be especially useful for evaluating drugs that are species-specific (e.g., leukotriene antagonists and 5-lipoxygenase inhibitors). In addition, although functional tests are not routinely used in this model, incorporation of measures of mobility and joint function may enhance the predictive value of the model.

Naturally occurring arthritis or autoimmune response:

MRL/lpr mice, Biozzi H mice and DBA/1 mice have been used to examine the onset of drug-induced tolerance and immunosuppressant drug effects on autoimmunity. The MRL/lpr mouse model has been useful for evaluating immunosuppressants and hormones.

Rat carrageenin-induced acute model of inflammation:

This model has been useful in assessing anti-inflammatory activity of cyclooxygenase inhibitors. Most of the animal models that involve inflammation in the paw may be used for measuring antiphlogistic action of a drug.

Adjuvant-induced arthritis in rats (AA):

AA in rats has been frequently used for screening nonsteroidal anti-inflammatory drugs and inhibitors of inflammatory cytokines as well as antimetabolite-like immunosuppressants.

Streptococcal cell wall-induced arthritis:

This model has been used for developing cytokine inhibitors.

Experimental organ transplant in animals:

This model has been used to identify the activity of immunosuppressants and antimetabolites, particularly those directed at cytolytic cellular immune processes.

Transgenic animal models:

A number of transgenic animal models are being developed for the study of rheumatoid arthritis and may prove useful over the next decade. Some examples include transgenic mice that carry genes for the env-Px region of the human T cell leukemia virus type I genome, human TNF, CD4, and HLA B-27.

3. Toxicology

Preclinical toxicology studies of a drug or biological product are designed to characterize general and specific toxicity using dosing routes and regimens as similar as possible to the proposed clinical trials with consideration of the demographics and disease status of the intended patient population. For instance, the prevalence of RA is high in females. Therefore, reproductive toxicity studies should be completed early in clinical development to support the inclusion of women of childbearing age in early phases of clinical trials. The need for reproductive studies for biological products is likely to be case-specific due to complications arising from immunogenicity and species selectivity. Therefore, standardized study designs, such as those recommended in the ICH reproductive toxicology guidance, may not be feasible or clinically relevant for biologics (ICH S5 1994). The need, and specific designs, for these studies may be discussed with Agency review staff.

Immunomodulatory or immunosuppressive agents administered to RA patients as monotherapy or in combination raise concerns about the adverse effects of prolonged immunosuppression. For example, malignancies (i.e., lymphomas) are a known risk of long-term, nonselective immunosuppression used for treatment of graft recipients. Investigational drug-related opportunistic infections and mortality related to immunosuppression have occurred in RA patients. Sponsors are encouraged to identify and use animal models that assist in selecting drug candidates that selectively inhibit cells and processes responsible for RA.

Antirheumatic drugs are often used in combination in an attempt to improve outcomes and minimize toxicities. However, drug interactions may result in increased toxicity, even at lower than previously evaluated doses of either agent. This concern is especially evident for agents that have long half-lives or

nonselective activity, or for drugs that share common target organ toxicity. Preclinical toxicity studies that evaluate the use of combined agents may be helpful in predicting clinical safety hazards. The duration of toxicity dosing of animals is usually linked to patient dosing regimens. Development and validation of in vitro or whole animal models is encouraged to address concerns regarding short- or long-term toxicity and to identify surrogate markers for patient immunocompetence.

B. Pharmacokinetic/Pharmacodynamic Strategies

In vivo pharmacokinetic studies should be used to evaluate drug disposition and metabolism, degree of linearity and accumulation, dose proportionality, and, for oral dosage forms, food interactions (Peck 1992). Some of these data can be gathered in a single study designed to evaluate a number of parameters. During formulation development, bioequivalence studies linking formulations may be recommended.

A particular concern with biological agents is the development of antibodies that may accelerate drug clearance or alter its distribution, resulting in changes in therapeutic benefit over time, or following repeated courses of treatment. To address this consideration, it is desirable for sponsors to build into their repeat-dose clinical protocols a coordinated evaluation of drug levels, receptor saturation, antidrug antibodies, and clinical responses. Optimally, these assessments would be conducted at the initiation of therapy and at several time points over the course of therapy. The presence of antidrug antibodies and their role in altering drug exposure, clinical activity endpoints, or adverse events would be evaluated. The goal of an integrated analysis of these parameters is to provide data to guide drug dosage or schedule changes to optimize therapeutic benefit. The best time for conducting these pharmacokinetic studies is prior to phase 3, before commitments have been made regarding dose and schedule.

Because polypharmacy is common during the treatment of rheumatic disorders, in vitro binding studies with blood from patients with active disease should be used as a preliminary screening tool for potential displacement reactions.

For products that may interact with rheumatoid factors (e.g., monoclonal antibodies), the frequency of patients with rheumatoid factor reactive to the antibody, as well as the effect of interactions on the pharmacokinetics of the product, should be evaluated.

C. Considerations in Phase 1 Trials

For general information on clinical development pertaining to most drugs and biological products, see the CDER guidance *General Considerations for the Clinical Evaluation of Drugs* (FDA 1978).

The term *phase 1* has two connotations: one refers to the earliest, first-time-into-humans trials, while the other encompasses studies of pharmacokinetics, metabolism, drug

interactions, special populations, and the other clinical pharmacology trials as described above. It is expected that both kinds of phase 1 trials ordinarily will be conducted during the clinical evaluation of therapies for RA. This section is primarily intended to discuss issues related to the first time people are exposed to the drug (including to a particular dose level, combination, or duration of therapy).

1. Settings and investigators

First-time-into-humans phase 1 studies should be carried out in institutions with a full range of clinical and laboratory facilities, and the patients should be kept under close observation. It is desirable that the trials be under the direction of physicians with experience in early drug development and rheumatology, or that a team of investigators combining experience in rheumatology and clinical pharmacology be employed.

2. Subjects

First-time-into-humans drug trials are frequently conducted in healthy volunteers. Such studies are predicated upon the ability to perform and to interpret the results of preclinical animal tests. If the preclinical testing does not reveal potential mutagenic or immune system effects, or potentially serious effects at or near the expected therapeutic range, testing in volunteers may be initiated. However, for products that have potentially serious toxicities, it may be appropriate for initial testing to be performed in patients with some potential to benefit. This has created challenges in selecting an appropriate initial patient population.

For products that have been tested in relevant preclinical toxicity evaluations and have been found relatively safe (without the potential for mutagenic, immune system or other serious effects at the proposed doses), trials may be initiated in healthy volunteers. However, if significant effects have been demonstrated or might be possible, selection of an appropriate patient population is necessary. It is recommended that patients meet the ACR criteria for both diagnosis and activity of RA and be without other serious medical conditions. Patients with minimal disease are sometimes not appropriate for the same reasons that the testing is not initiated in healthy volunteers. Patients with devastating RA may also not be the best starting population because of the medical complications of their disease. In addition, they may be less likely to respond to therapy.

There is ongoing epidemiologic work on identifying markers of increased risk in RA. These could be useful for identifying patients with poor prognoses, who might be considered for very aggressive treatments of potential high toxicity (e.g., immunoablative therapies followed by stem cell transplants). Application of epidemiologic studies may allow a very aggressive treatment to be restricted to a subset of RA patients who have a demonstrated shortened life span due to their

disease (e.g., subjects with greater than 30 affected joints or a HAQ score with fewer than 75 percent of questions answered "without difficulty").

When the characteristics of the agent suggest that it may potentially have long-term gonadal effects, it is desirable that men and women not wishing to parent children be chosen for phase 1 studies.

3. Trial design

Ordinarily, initial phase 1 studies are sequential dose escalation trials, in which safety and tolerance at a specific dose is established before exposing additional subjects to a higher dose. A single dose is almost always tested first, followed by repeated dose studies; however, this design is influenced by the type of agent used. Although escalating the dosage to a clearly determined maximum-tolerated-dose (MTD) will aid future trial design, in some instances it is not medically prudent to try to fully characterize the MTD. Additionally, for some products, an MTD may be undefinable.

The starting dose chosen is often a *no-adverse-effect* dose (determined by interspecies milligrams (mg)/meter square/day dose conversion from animal to human). For biologics, the initial dose chosen is often one thought to have no adverse biological effect, with caution regarding the possibility of relative species specificity and comparing receptor avidity between test species and humans. Conservative dose escalations (e.g., half log or less) are usually recommended.

4. Concomitant therapy

Use of low-dose corticosteroids (up to 10 mg prednisone equivalent daily) and NSAIDS may ordinarily be continued in phase 1 trials. Concomitant therapy with methotrexate and similar agents should be avoided in initial phase 1 trials of all novel antirheumatic drugs, biologics, and devices because of the difficulty of differentiating the toxicity of the novel agent from that of the co-administered product.

Physicians now prescribe methotrexate and similar agents earlier in the course of rheumatoid arthritis. Recruiting adequate numbers of patients not taking these agents may be difficult. Approaches that may allow the use of methotrexate and similar agents in later phase 1 trials include (a) obtaining reassuring evidence of lack of toxicity from relevant animal models in which co-administration occurred and (b) starting at doses significantly lower than the no-adverse-effect level of the single agent as determined by earlier phase 1 studies or preclinical studies, or both. Such proposals should be discussed in the planning stages with Agency staff.

5. Observations

a. Safety

The standard batteries of safety observations have been described elsewhere (ICH *S5A* 1994). However, additional types of safety observations may be necessary (e.g., tests of effects on cellular and humoral immune function or host defenses). For products with the potential for effects lasting long after administration, or for delayed toxicity, appropriate follow-up should be designed. For example, phase 1 studies of agents used to deplete or modify the function of T-cell subsets should be designed to carefully assess both the short- and long-term effects on number and functional status (e.g., delayed type hypersensitivity responses) of cell populations and other pertinent pharmacodynamic assays during therapy and during follow-up.

It is also desirable to incorporate individual patient adverse event stopping/withdrawal *rules* into protocol designs. In addition, it is often advisable to incorporate into trial designs rules for trial stopping or trial modification in case adverse events are observed. For example, dose escalation rules should be clearly defined in dose-finding studies, with provisions for enrollment of additional patients at or below the dose-causing toxicity if possible significant adverse events are observed.

It is desirable to develop a standardized toxicity grading scale for use in all trials of a product based on the known and suspected toxicities of the product or of the drug class. This may improve consistency of adverse event reporting and allow more accurate comparisons among trials.

b. Efficacy

Developing an understanding of the agent's therapeutic potential in early trials is highly desirable for efficient product development. This may be attempted in phase 1, but can ordinarily be achieved only by performing controlled trials. RA responses in open trials are of little value in indicating efficacy. Consideration should be given to the more modest goal of determining whether the pharmacological effect predicted from the preclinical development is present (proof of concept).

D. Considerations in Phase 2 Trials

During phase 2, larger, often longer, term trials are employed to better define the dose- and exposure-related activity and toxicity of the agent. Enough information should be generated to ensure that the phase 3 trials can be conducted safely and with a probability of success. In addition, phase 2 trials should solidify a total drug development strategy to ensure that, after the phase 3 safety/efficacy trials are done, all of the information needed for registration will have been gathered, including an appropriate safety database, clinical

pharmacology, dose-response data, the exploration in special populations (e.g., renal failure, hepatic failure, pediatric patients), and information on drug interaction with agents expected to be co-administered.

There is nothing to preclude conducting additional phase 1 clinical pharmacology studies and phase 2 trials while the phase 3 development is ongoing.

The following issues are important for phase 2 trials in RA:

1. Trial design

Dose finding is a central challenge of phase 2 development. Once a reasonably safe range of doses has been established, randomized, parallel-arm dose-comparison trials are ordinarily recommended. The use of a placebo arm is desirable for several reasons. First, if no difference is found among doses, there is usually no other way to determine whether all doses were equally effective or equally ineffective. Second, if a dose-response trend is found, the placebo arm may indicate the possible magnitude of the observed effect. If use of a placebo is impossible, designs should include wide dose ranges or durations, or repetitions. Active controlled designs that specify an arm with a well-characterized, known therapy can also be very useful.

Signs and symptoms measures may be used for dose-finding studies, but it is possible that separate dose-finding studies may be needed for longer term endpoints.

For agents that are thought to have prompt onset and rapid offset of effect, alternative designs, including crossover and titration designs, may be useful, although historically this has not been the case. Trials of two or more doses that permit liberal titrating per the patients' responses are unlikely to clearly demonstrate a dose response because these titration designs result in a blurring of any existing dose distinction.

The desirability of identifying a range of doses with acceptable toxicity and reasonable activity for study in phase 3 cannot be stressed enough.

2. Safety

Every RA investigational therapy raises safety concerns. Whenever there is a potential for significant toxicities, long-lasting or delayed-onset, it is desirable to design the phase 2 studies to provide a group of patients with longer term follow-up preceding the larger phase 3 studies. Provisions for long-term follow-up can be helpful in addressing issues raised during premarket review (e.g., potential for immunosuppression, opportunistic infections, neoplasia, and induction of

autoimmune disease). Standard toxicity grading scales and stopping rules are also desirable in phase 2.

3. Additional development aspects

a. Concomitant therapy

Before starting phase 3 trials, an evaluation of the test product's interaction with other agents likely to be used by the target population should be performed. Initial information can be established based on metabolic pathways, studies of in vitro systems, animal or human pharmacology studies, or drug interaction studies. This type of information helps in directing areas in need of clinical evaluation. When products are intended to be tested as combination therapy with the investigational agent, substantial information on interactions and safety of co-administration should be developed in phase 2.

b. Gender effects

Most RA trials have predominantly female enrollment. Sponsors should evaluate whether the observed safety and efficacy findings are restricted to women or can be also extrapolated to male subjects. This may be accomplished by subset analyses from trials, PK data, or other information (FDA 1993).

E. Efficacy Trial Considerations

The overall goal of phase 3 work is to demonstrate the efficacy of the product in convincing controlled trials and to accrue a sufficient safety database. Efficacy trial protocols should contain an analytical plan that precisely identifies the primary comparison(s) to be made, the criteria for success of the trial, and the statistical tests that will be used. Protocols should be designed to generate data that support desired labeling claims. Any additional planned, ongoing, or completed trials that are also intended to generate data supporting the claim should be identified.

1. Global considerations

a. Patient selection

Activity of disease: Unless some other specific subgroup is targeted, patients enrolled in efficacy trials should at a minimum meet the disease definition and disease activity as defined by ACR criteria. Consultation with the Agency on the generalizability of claims derived from trials with significant limitations on entry criteria is recommended.

To enhance the power of the trial, strategies to improve the chances of a response to therapy are often employed. Some designs incorporate an attempt to select active patients by withdrawing background treatment and allowing patients to *flare*. Only individuals with sufficiently high scores are enrolled. The relevance of this type of observed flare is questionable, and its ability to predict the normal course of active disease has not been established. Many patients randomized to placebo in such studies exhibit the characteristic response of rapidly returning almost to baseline without further treatment. In addition, when patients undergo blinded withdrawal from therapy within these trials, similar dramatic flares are not observed. This raises the question of whether there is an expectation bias on the part of patients, who have been told about the flare procedure, and ascertainment bias on the part of investigators, who wish to have patients meet the entry criteria and enroll in the study. These uncertainties and instabilities around the outcome measures used in such trials should be kept in mind when employing these designs.

A proportionately smaller, but nevertheless noticeable and prompt, *regression to the mean* is noted in the joint scores of patients required to have a certain minimum value for trial entry in trials not employing a *flare strategy*. This means that patients, on the whole, will not actually have disease as active as anticipated when the entry criteria are set.

Subgrouping patients by disease markers: RA is likely composed of a number of more or less distinct diseases delineated by a common genetic background, corresponding clinical manifestations, similar serologies, and responses to therapy and prognoses. The study of RA possibly may be made more efficient with the use of markers with clear prognostic significance as entry criteria to increase patient homogeneity. Novel epidemiologic and molecular genetic approaches may lead to identification of even more subgroups. However, prospective studies are first needed to confirm the clinical usefulness of new purported prognostic factors. Where existing data do support markers as prognostic indicators (risk factors), such as the presence of rheumatoid factor, erosive or vasculitic disease, and DR4 homozygosity, they can be taken into consideration in the design of trials, as can factors known to affect treatment responses. Although in some cases such studies could limit generalizability and impact labeling of the final product, it is also possible that such targeting could improve the risk/benefit profile.

b. Concomitant antirheumatic therapy

Studies in RA patients, except in those with very mild disease, are carried out in the presence of concurrent active therapies, including steroids, NSAIDS, hydroxychloroquine, etc. This concurrent therapy creates

numerous challenges in patient selection, toxicity monitoring, and clinical trial design. For example, since methotrexate therapy is used to treat many RA patients, it is inevitable that new agents will be used in combination with methotrexate in clinical practice unless a contraindication exists. Therefore, unless a prohibition on concurrent methotrexate is supportable, data regarding use of the investigational agent in combination with methotrexate are needed to evaluate the potential for immunosuppression from combination therapy. Other agents should be similarly evaluated.

In addition, patients can be categorized according to their prior responses to standard therapy. Varying trial designs may help assess the response of different response categories to an investigational therapy. For example, with respect to methotrexate use, the RA population can be divided into five groups: (1) methotrexate noncandidates — disease too mild or too early for methotrexate; (2) methotrexate candidates — disease sufficiently (or will become sufficiently) active to justify methotrexate; (3) methotrexate successes — disease reduced to negligible amounts; (4) methotrexate failures — clear drug failures, for inefficacy or tolerability, and (5) methotrexate *partial responders* — with considerable residual disease despite methotrexate. Each of these groups might be considered separately for candidacy for an investigational agent and with respect to an appropriate trial design. If only a subpopulation of RA patients (e.g. methotrexate nonresponders) is studied in a particular trial, the results strictly reflect efficacy in that group only, but they may, of course, imply something about efficacy in other groups. Single trials in various responder subpopulations could be supported by positive results in other subpopulations. Any planned subpopulations should be clinically distinguishable. Sponsors should consult Agency personnel when planning a clinical development program contemplating an RA claim that is limited to a subpopulation with the disease.

c. Other concomitant therapies

Most patients with RA are taking concomitant medications. Use of medicines unlikely to influence treatment outcomes (e.g., antihypertensives) should simply be recorded, although investigators should be alert for possible drug interactions. Obtaining information during clinical development on co-administration of the test medication and expected concomitant medications is desirable. The following approaches may be considered in dealing with arthritis medications or analgesics.

Prohibit use: This strategy may result in noncompliance or an increased number of dropouts.

Incorporate protocol-specified use, with monitoring: With this strategy, additional analgesic use (and possibly other arthritis medications) may be used according to protocol-specified criteria. In addition, for long duration studies, protocols should address (1) whether intra-articular steroids are permitted and, if so, for how long assessments of the involved joint are excluded from analysis; (2) the manner in which *stress* doses of corticosteroids for surgery, etc., are to be handled; and (3) how soon after such doses protocol assessments would be allowed.

Design analgesic use, or its quantitative consumption, as (part of) an efficacy endpoint.

Define use of more arthritis treatments as (part of) an efficacy endpoint.

d. Stratification

Randomization is intended to balance populations for confounding variables; however, there is always a chance that randomization may fail to achieve balance, particularly in smaller trials. It may be advisable to stratify known (or highly suspected) major risk factors to ensure their balance across arms. Any factor whose influence on the outcome is suspected to be as strong as the treatment's influence should be considered for stratification (e.g., erosive disease, presence of rheumatoid factor). An often overlooked risk factor is the patient's past therapeutic history. (See statistical section for further discussion.)

e. Blinding

Because most RA outcome measures have a high degree of subjectivity, the highest confidentiality in patient and assessor blinding should be sought to achieve a credible inference. Blinding may be compromised if there is not an approximate parallelism in time to onset, nature of response, and toxicity profile between active and controlled interactions. Trials should have parallel (e.g., "double dummy") dosing in all arms when possible so that a drug requiring frequent dose manipulations does not threaten the blind. If *arm specific* treatment adjustments are necessary (e.g., per monitored drug levels), these can be done by an unblinded (and sequestered) third party to maintain patient and assessor blinding. In this case, parallel changes should be made as dummy adjustments in the control arm to preserve blinding. Similarly, if the blind is likely to be compromised by infusion-related events or other features of the treatment protocol, critical treatment endpoints, such as joint counts, should be assessed by an independent party with no knowledge of the subject's history.

f. Effects of dropouts and noncompliance.

It is important that trials be designed to minimize dropouts and the attendant information loss. Traditionally, recommended RA trial designs have focused on eliminating sources of variability, for example, extra pain medications, and intra-articular injections. Often, these interventions were defined as major protocol violations, requiring that the patient be dropped from the study. There is a trade-off between patient retention and tolerance of variability in RA trial design. Protocols demanding rigid adherence may yield uninterpretable results because of dropouts and noncompliance emanating from patient and investigator intolerance of the requirements. On the other hand, protocols permitting any kind of additional intervention may likewise be so confounded as to defy interpretation.

The following strategies may help minimize loss of information:

i. Use screening or run-in periods so that patients are randomized to treatment groups only after their eligibility and commitment are confirmed.

ii. Thoroughly train investigators and study personnel to minimize inappropriate enrollments, protocol violations, and other deviations that would decrease the ability to assess trial outcomes.

iii. Include dropouts in the definition of the endpoint, as in a time to defined treatment failure, or a defined by-patient success or failure. A sufficiently late time point should be chosen as the endpoint to avoid the situation where the time course for response differs between two active therapies in a trial, which can introduce bias.

One example of this approach would be to use a protocol-defined response rate as the primary endpoint, wherein dropouts due to lack of efficacy are classified as *nonresponders*. With this type of endpoint, the criteria for classification as a nonresponder should be clearly and prospectively defined. The use of this type of endpoint could be justified in situations where there are robust phase 2 data suggesting drug responsiveness at a defined point after initiation of therapy and durability of that response. In this case, one could define the primary analysis as a comparison of the proportion of patients with an ACR 20 response at six months. The protocol could specify that if no improvement compared to baseline were seen on two consecutive study visits after two months on protocol, the subject would be declared a nonresponder. Experience will determine whether this approach effectively limits information loss due to dropouts.

iv. Make provisions for following patients who have stopped experimental treatment. Options include allowing a protocol-specified crossover to a standard therapy for patients meeting predefined criteria for treatment failure.

v. Allow more flexibility in treatment options during the study. Some designs that have been used include allowing dose adjustment of the comparator arm (assessor and patient blinded), allowing add-on therapy for patients meeting predefined criteria for inadequate response, and allowing a limited number of joint injections, with elimination of that joint from assessment.

2. Trial designs in RA

Clinical trials in RA can be designed in a variety of ways. More than one claim can be pursued in the same trial, and claims can be submitted singly or together. Trials can be designed to test a difference — demonstrating that the investigational product is superior to control (placebo, lower test dose, another active agent), or they can be designed to test no difference — demonstrating that the product is adequately similar in efficacy to active control. Placebo-, dose-, concentration- or active-controlled designs can be used.

Because the persuasiveness of trials showing a difference is, in general, much greater than that of equivalence trials, it is highly desirable for a claim to be convincingly demonstrated in at least one trial showing superiority of the test agent over placebo or active control. If a claim of superiority over a specific comparator is sought, rather than just straightforward efficacy, the claim should be substantiated by two adequate and well-controlled trials showing superiority. Such trials can also be the basis for demonstrating the product's efficacy.

a. Superiority trials

The standard two-arm, investigational agent versus placebo design has been the most common RA design and is the most straightforward. The details of trial design will depend on the population tested. Patients with mildly active RA taking only NSAIDS, who have never been treated with an additional class of therapy, may be enrolled in a placebo-controlled trial with continuation of NSAID background therapy; however, patients doing poorly on NSAIDs alone are usually not appropriate candidates for placebo-controlled trials. The same considerations apply to patients who are partial responders to, or who have failed, various other treatments.

Alternatives to the two-arm difference design are a standard dose-response study and a superior-to-active-control hypothesis. These designs may

accommodate the need to provide active treatment to patient groups where randomization to placebo is infeasible.

b. Equivalence trials

Equivalence trials are designed to demonstrate that the test drug is *adequately similar* to an active control. This is done using a prospectively defined *equivalence test*, specifying to a 95 percent confidence level that the real difference between test and control is smaller than some predetermined amount. Achieving similar point estimates of efficacy of two agents is ***not*** a demonstration of equivalence.

Equivalence trials can serve two purposes. First, they can be used to supply evidence for a simple efficacy claim. Second, they can be used to pursue a specific *equivalence to drug X* claim. Both purposes could also be pursued in the same trial. The important point to note is that ***the strength of the evidence may need to be stronger for a claim of equivalence to drug X than for a simple efficacy claim.*** Thus, the equivalence test may differ, depending on which claim is intended. Currently, the equivalence standard that is appropriate for a given trial in RA will be determined on a case-by-case basis. As noted above, this test may be more stringent if a claim of *equivalence to drug X* is being pursued. Additionally, the test of equivalence may be constructed differently if a placebo arm is present, since the presence of the placebo allows estimates of absolute and relative drug effect size.

In either case, the statistical test for equivalency needs to be quantitatively described in the protocol. Under either the pursuit of a simple efficacy claim or the pursuit of a specific *equivalence to drug X* claim, the basis of the decision on an appropriate test remains, fundamentally, a clinical one. It represents a consensus, in that particular circumstance and for that particular claim, on what small potential difference can be considered clinically insignificant, to allow the treatments to be considered clinically equivalent.

There is considerable experience in the interpretation of active-controlled trials in clinical situations where the response to the intervention is high. As an example, antibiotics are evaluated by the Division of Anti-Infective Drug Products (CDER, FDA). For these products, the magnitude of the potential difference permitted in an equivalence determination depends on the response rate of the standard treatment. For example, a new agent being compared to standards with response rates of 90 percent or more will be allowed a 10 percent window to provide confidence that the difference between the standard response rate and the test response rate is no more than 10 percent. Technically, this means the 95 percent confidence interval

of the difference must lie fully to the right of 10 percent. If the standard treatment is known to have an 80 to 90 percent response rate, a 15 percent window is used. These equivalence tests were designed for trials without a placebo arm and for clinical situations where the placebo response is known to be very low.

Treatment response rates in RA are often in the vicinity of only 50 percent (depending, of course, on the endpoint used) with placebo rates of about 20 percent, so the clinical decision for an allowable small difference may differ from that with antibiotic therapy. At this time, the decision will have to be individualized for each trial.

A major problem in equivalency trials lacking a placebo arm is ensuring that both treatments are equally effective, rather than equally ineffective. A number of the agents that are approved for RA have fairly small effects and may fail to show efficacy when tested against a placebo. Comparative trials intended to show *equivalence* to such treatments, when not anchored by a placebo control group, may lack credibility. Thus, it is desirable in equivalence designs to select highly effective comparative agents. If possible, use of a third (placebo or lower dose) arm, so that a treatment difference can be shown, is a desirable strategy in equivalence trials. This arm would not necessarily have as many patients or as long a duration as the active comparators. If a placebo arm is present, both the test and active arms need to statistically exceed placebo for a finding of *equivalence* to have meaning.

Strict attention to numerous aspects of trial design and conduct are important to ensure accurate inferences from equivalency trials. Design decisions regarding patient population, dosing, and efficacy and safety assessments should be done in a manner that is unbiased against the control to ensure a *fair comparison*. Furthermore, attention to certain problems in trial conduct, such as minimizing dropouts, noncompliance, and missing data is essential to the reliability of the inference. These aspects of trial conduct may obscure differences and lead to a false conclusion of equivalence. This is the opposite of their effect in a difference design to show superiority, where they work against trial success.

In any particular development, the choice of trial design depends on many factors. Since controlled evidence showing a difference is more persuasive than that showing equivalence, greater efficiency (fewer patients or shorter exposures) is available with development strategies using trials employing maximal differences between trial arms. Optimally, this means placebo controls, with the requisite *background therapy*, given to all patients.

c. Trial designs novel to the study of RA

Although not used traditionally in the study of new RA treatments, the withdrawal design can be considered in certain circumstances. The withdrawal design is sometimes used to assess efficacy. In this design, patients in both arms of a study are treated with the investigational agent, which is then blindly withdrawn from one arm, after which patient outcomes are compared. Showing that patients taken off the investigational drug get worse demonstrates effectiveness. Natural endpoints for withdrawal designs are *time to (predefined) worsening* using standard *time-to-occurrence* statistical tests or a simple comparison of proportion of outcomes in the two arms. Withdrawal studies may be performed with both arms on background therapy.

There are a number of caveats about withdrawal designs. If the product is very toxic, so that only a small (tolerant) subset of the original population remains at the end of the trial and is available for the double-blind withdrawal phase, the generalization of any inference from the withdrawal design is limited to that tolerant subset. Additionally, it should be noted that, if a drug induces habituation or tolerance, withdrawal or rebound phenomena may make withdrawn patients worse even though drug therapy did not have a beneficial effect.

3. Analytical Issues

a. Handling dropouts

Historically, RA trials have suffered from information loss due to dropouts. Dropouts probably never occur randomly, and rarely occur fully independent of the treatment being tested, so there is always the possibility that dropouts introduce a bias. This problem is common in many randomized trials. Methods for analyzing the effects of dropouts have been proposed, but none is fully adequate.

The problem of dropouts is not resolved by using an intent-to-treat (i.e., all randomized patients included) analysis with an imputation by last-observation-carried-forward (ITT/LOCF) or by showing that both the ITT/LOCF and PP/OC (per protocol completers/observed cases only) analyses concur, although these approaches may increase confidence in the results. It should also be noted that there are other methods of modeling missing data, for example, see Little and Rubin (1987). Such modeling methods require assumptions that are nonverifiable by existing data.

The effects of dropouts should be addressed in all trial analyses to demonstrate that the conclusion is robust. One trial design approach is following all patients, including dropouts, to the planned trial endpoint, even if postdropout information is confounded by new therapy, and

performing an analysis including these patients. Another approach involves the *worst case rule*: assigning the best possible score to all postdropout placebo patients and the worst score to all postdropout treatment patients, then performing an analysis including these scores.

b. Comparison to baseline outcome measures

A phenomenon frequently observed in RA, as well as in other conditions, is that patients who stay in trials do better than those who drop out: responders do better than nonresponders. This is true both for placebo groups and active treatment groups. If observations of the disease were made exclusively from clinical trials, one might conclude that the natural history of the disease is inexorable improvement. This phenomenon is attributable to preferential dropout of worsening patients (a phenomenon not adequately compensated for in LOCF analysis) as well as *regression to the mean*. The problem is exacerbated in flare designs, where all patients have major improvement regardless of treatment status. This fact makes comparison-to-baseline outcome measures difficult to assess, since, very often, much of the improvement noted has no relationship to a treatment effect. For these reasons, active-controlled trials not incorporating a placebo arm and using comparisons to baseline may be extremely difficult to interpret, especially if a flare design is employed. In any case, success in any trial implies improvement over control.

4. Statistical Considerations in Efficacy Trial Design

It is advisable to discuss the design and analysis with the FDA review team prior to embarking on a study. In addition, FDA's *Guideline for Format and Content of the Clinical and Statistical Sections of New Drug Applications* (1988) contains useful information.

a. Randomization/stratification

Randomization is intended to allocate patients to treatment groups to avoid bias and to ensure that statistical procedures can be appropriately applied.

In some clinical trials, there are known factors that are at least as influential in controlling the observed severity of disease as the drugs being studied. Stratification may be used to avoid relying on randomization properties to balance patient assignment for these factors. Stratification is implemented by constraining simple randomization to balance the assignment of patients to treatment groups for the chosen stratification factors.

Every phase 2 and phase 3 study protocol should contain a randomization section. All constraints imposed on the randomization should be explicitly

identified. It can then be inferred, when a stratification factor or sample size allocation constraint is not mentioned in a protocol, that there exists no corresponding randomization constraint. This applies to whether patients are blocked to balance treatment assignment for time of patient entry into study and to the more obvious stratifications on center and baseline.

Because stratification implies constraints on randomization, studies that have been stratified for certain factor(s) should account for these factors in the statistical analysis section. The protocol-defined analysis should be implemented for each study.

There are also statistical procedures to address bias in treatment group comparisons by adjusting for factors (covariates) that, like the stratification factors, are to be prespecified in the protocol or by using a mechanism to determine a fixed number of covariates prespecified. It is important to prospectively identify covariates (or criteria for choosing covariates) in the protocol. In addition, statistical adjustment procedures should be used with care and include appropriate verifications of assumptions.

In deciding whether to stratify randomization in all clinical trials, practical judgment is required. There are reasons to choose stratification and reasons to choose statistical adjustments.

The first advantage of stratification is that it avoids possibly major statistical adjustments of differential treatment effects. Stratification would essentially eliminate the effect of such adjustments before analysis began. Second, although stratification and statistical adjustment procedures are both designed to remove bias in estimated treatment effects, stratification is usually more powerful. This is because stratification leads to smaller variances of estimated treatment effects than does statistical adjustment without stratification. Finally, the inclusion of stratification factors into a statistical analysis model should result in increased power to detect effectiveness.

Stratification becomes increasingly clumsy as the number of strata increases and, consequently, the available number of randomizable patients per cell decreases. In this case, it is logistically simpler to not stratify, but to rely on statistical methods to adjust for these factors.

The best approach may be to combine stratification, applied to a limited number of the most influential prognostic factors, with statistical modeling. This protocol-defined statistical modeling would both account for stratification and be used to adjust for the effects of a parsimonious number of the most important remaining factors.

b. Identification of primary efficacy variables

Each phase 2 or phase 3 study protocol should identify the primary and secondary efficacy variables. Primary efficacy variables are critical to the identification of the effectiveness of the product. Secondary efficacy variables are those that support the validity of the primary variables but are less critical in deciding if this product is effective. It is helpful, but not necessary, that statistical evidence of efficacy be shown for secondary efficacy variables.

c. Prespecification of statistical analysis

Statistical analysis of clinical endpoints is part of the process for obtaining consistent and convincing evidence of product efficacy. These statistical analyses should not be data driven. This is implemented by identifying, in each study protocol, before data are available for analysis, a sufficient description of the statistical analyses of primary efficacy variables so that an independent statistician could perform the protocol analyses. A brief description of the statistical analyses should include but not necessarily be limited to specifying: (1) the level of significance to be used; (2) whether statistical tests of hypothesis or confidence intervals will be 1- or 2-sided; (3) whether interim analyses are planned and, if so, how the tests of hypotheses and confidence intervals will be adjusted to account for interim looks at the data; (4) the mathematical expression of the statistical model(s) used; (5) the minimal statistical results needed to demonstrate a successful outcome; (6) the treatment of missing values and dropouts; (7) the method used for controlling type I error rates for multiple primary efficacy variables; and (8) the method used for making multiple treatment comparisons.

d. Multiple endpoints

There has often been a clinical argument for using multiple endpoints to assess primary evidence of effectiveness in RA. The theoretical bases for such combination endpoints are the focus of an area of ongoing statistical research. For example, for the four measures recommended in FDA's previous guidance (FDA 1988), trial results were considered to support a conclusion of effectiveness when statistical evidence of efficacy was shown for at least three of the four measures: physician global assessment, patient global assessment, swollen joint count, and painful joint count.

Multivariate statistical methods are also available for analyzing the set of primary efficacy variables. Procedures are being developed for inferences derived from multiple endpoint results.

Efficacy variables can be combined within patients (composite endpoint). Such a fixed combination of efficacy measures should be well defined in the study protocol. Composite efficacy variables have the advantage of avoiding several statistical and inferential difficulties associated with multiple endpoints.

e. Dropouts

Dropouts are patients who, after a certain period of time in a trial, fail to provide clinical efficacy data scheduled to be collected by the protocol. Frequently, dropouts occur for reasons related to taking the assigned test drug (adverse effects or lack of efficacy). Since dropouts do not usually occur randomly, the remaining patients constitute a biased subsample of the patients originally randomized. Unless trial dropouts can be assumed to be random (which is rare), no model used to impute their effects can be verified.

Methods used to handle dropouts, such as LOCF and *completers* analyses, are not fully satisfactory even though they have often served as the basis for determining that adequate statistical evidence of efficacy has been provided. The LOCF method usually does not preserve the size of the test, either for the comparison of final observations or for the comparison of rates of change. Alternative methods include growth curve analysis and random effects regression. These are also susceptible to informative censoring — that is, dropping out depends on the value of the response. It is often useful to show that the results hold for a variety of analyses (i.e., they are robust).

f. Trials with several treatment groups/multiple comparisons

In clinical trials involving more than two treatment groups, a statistical multiple comparison procedure controlling the experiment-wise error rate to 5 percent or less should be applied. In essence, there should be overall statistical evidence of a treatment main effect before attempting to make specific drug comparisons relevant to proposed drug labeling.

g. Trials simultaneously used to pursue more than one claim

A single trial can be used to pursue simultaneously more than one claim; an adjustment of significance level for multiple analyses is not always necessary. If the order of testing the hypotheses is prespecified, then no penalty need be taken. For example, when a trial is simultaneously pursuing a six-month signs and symptoms claim and a twelve-month x-ray claim, if the trial *wins* by the first hypothesis tested — signs and symptoms

— then the x-ray hypothesis can be calculated without an adjustment penalty.

h. Interim analyses

Interim analyses are those that, for any purpose, are repeatedly performed on accumulating clinical trial efficacy data. Methods have been developed to compensate for the fact that multiple tests (including interim analyses) alter the true significance level. The study protocol should state whether such interim analyses are planned or not. Should interim analyses be planned, the plan and its implementation should be described in the protocol. The description should include who will have access to the interim data, the scheduling of these interim analyses, the method to be applied for adjusting significance levels, and the corresponding time sequence of significance levels at which statistically significant results will be claimed.

Although an interim analysis may not be thought to affect the subsequent collection of efficacy data, interim analyses carry the additional risk that the blinding or conduct of a study may have been compromised. Statistical methods cannot compensate for any unblinding and bias that may result from gathering the information needed to perform an interim analysis. Finally, if any major protocol change becomes necessary (e.g., a new therapy becomes available), it is important that such a change not be influenced by those unblinded to data.

i. Sample size

Failure to recruit an adequate number of patients is a major reason why an effective product may fail to meet established statistical criteria for efficacy, independent of whether the purpose was to show superiority or comparability of treatment effect. The method for determining the sample size should be stipulated in sufficient detail to permit independent verification of the computation. This should include identifying the efficacy variable that the sample size determination is based on, the magnitude of the hypothesized clinical difference, the standard deviation, the power, the significance level, and the sidedness of the statistical procedure(s) described in the analysis plan. Furthermore, the size of the clinical difference chosen should be justified, and the rationale for the choice of the efficacy variable used to determine sample size should be discussed. For comparability from one trial to the next, it is optimal to use the same efficacy variables as were used to power earlier studies.

j. Trials to show clinical equivalence

The words *clinical equivalence* are used in a much narrower sense than these words might imply to the casual reader. First, there is often no intent to show equivalence of two or more drugs across the broad spectrum of pharmacologic effect. Rather, focus is on showing no clinically relevant differences for one or possibly more variables that are to be clearly identified in advance. The concept of equivalence is two-sided in that if, for any outcome measure, one drug is sufficiently different from another drug, then these drugs are no longer deemed equivalent in that variable.

To show equivalence, the variables serving to measure these effects of interest should be defined in the protocol. For each efficacy variable for which clinical equivalence of effect is sought, the magnitude of a difference deemed to be inconsequential should be identified. The clinical data should then show, with 95 percent confidence, that this predefined difference is not exceeded.

Inference based on trials to show equivalence is inherently less convincing than inference based on trials to show the existence of a difference. Often, clinical trials do not detect treatment differences that are known to exist. In such cases, statistical methods may then seemingly provide evidence of equivalent effect (e.g., to placebo).

k. Appropriateness of the statistical methodology

The appropriateness of the statistical model should be assessed, including checking for outliers and determining if distributional assumptions (usually normality) are met and if common variance assumptions hold.

l. Site effects

If the patients have been stratified and randomized by site, the analysis should include a site effect. There may be a site-by-treatment interaction reflecting the degree to which the treatment varies across sites. This is often notable when there is a great variation in enrolled patients across sites. Site-by-treatment interaction should be explored.

F. Safety Analysis

The approach to evaluating adverse event data and laboratory values has traditionally differed from that used to evaluate efficacy. The purpose of safety evaluations is usually not to test a specific hypothesis, but to examine the pattern of effects and to detect unusual or delayed events. Analyses using cumulative occurrences, scatter-plots of laboratory values (baseline versus on-therapy), or other techniques may be helpful. The safety profile should address to what extent adverse events (drug reactions or lab values) depend on duration of drug exposure, dose level, coexisting medical conditions, or

possible drug interactions. Incidence rates should be calculated using denominators that reflect the period of drug exposure for the population at risk. Cumulative incidences (hazard rates, instant probabilities) do a better job representing the temporal pattern of drug effects than do prevalence rates; comparative cumulative incidence tables — drug versus active control(s), versus placebo — also are very helpful to practitioners. Critical incidence rates should be described with 95 percent confidence intervals.

Ensuring safety during the development of a drug, biologic, or device can be optimized by adequate preclinical evaluations and the development of a standardized clinical safety assessment system. Elements of a successful safety assessment system include the use of predefined standard terminology (such as the *Medical Dictionary for Regulatory Activities Terminology*) and criteria to define and assess adverse events (AEs), approaches to optimize AE detection, and appropriate safety stopping rules in trials. It is also useful to capture AE severity (grades 1 [mild], 2 [moderate], 3 [severe], or 4 [life-threatening]), outcomes (such as the need for therapy and whether resolution or death occurred), treating physician assessment of association with study agent (remotely, possibly, or probably related), and impact on the trial (none, dose of agent delayed or changed, or patient withdrawn from further therapy). Stopping rules, determined by the risk/benefit ratio for the agent in the study population, are desirable both for individual patients (a single grade 3-4 AE is often used) as well as for the clinical trial, especially in dose-escalation studies.

1. Intrinsic trial design considerations

An attempt should be made to characterize the patient population susceptible to adverse drug effects. Some extraneous factors, such as variations in soliciting and reporting adverse events among the investigators and differences in the definition of normal ranges for lab values among different laboratories can complicate the safety data. Since adjustment for their effects may be difficult, precautions should be taken in the design stage of the trial to minimize the influence of these factors by preparing clear and specific instructions for data collection and monitoring adherence of the investigators and the laboratories to the protocol. Procedures for normalizing laboratory data, for example, may be employed. As previously mentioned, developing standardized toxicity grading scales that can be employed in all studies may also be useful.

2. Adequate numbers

The ability to detect adverse experiences is dependent on the number of patients evaluated in the clinical trials and in clinical usage. In studies of 300 or more patients having adequate exposure to the investigational drug, it is expected (with 95 percent confidence) that at least one patient will manifest an adverse event having an incidence rate of 1 percent or greater. Smaller studies fail to meet even this minimal incidence detection standard. In most cases, however, it is possible to combine studies of equal duration to establish adverse experience rates.

For any chronically administered product, the safety database should include at least 300 patients treated with the maximally recommended dose for at least six months and at least 100 patients treated for at least twelve months (ICH *E1A* 1995). Larger and/or longer safety databases may be advisable for agents with known or potential safety problems.

IV. SPECIAL CONSIDERATIONS FOR BIOLOGICAL PRODUCTS

Although there are similarities between RA trial designs for drugs and biologics, biologics have special characteristics and problems that should be considered in their development.

A. Species Specificity

The schemes used traditionally in determining the initial human dose may not pertain to biologics. Biological agents may behave differently in animal models than in humans, depending on the physiologic relevance and avidity for the receptor of the ligand in the animal compared to the human. Immunogenicity may also be species specific.

B. Dose Responses

The dose-response curve may be steep and/or even hyperbolic, and an agent can be quite toxic at levels just above those thought to show efficacy.

C. Toxicity Response

The toxicity response may be highly unpredictable and potentially very dangerous and may include the risk of disease worsening. Agents may have narrow therapeutic windows. Biologics have the potential for disruption of immunologic and physiologic processes. Monoclonal antibodies to cellular epitopes of the immune system, for example, or to TNF receptors may cause serious morbidity at doses only slightly higher than those that are efficacious with markedly less toxicity.

D. Product Homogeneity

This often plays a critical role in activity and toxicity of a compound. Product alterations can greatly affect physiologic activity. Thus, biologics should demonstrate lot-to-lot consistency to the extent possible while under development and be reasonably well characterized to be properly evaluated.

E. The Role of Antibodies

If phase 2 data suggest that agent-induced neutralizing antibodies could interfere with the efficacy of a biological agent over time, it may become necessary to formally investigate

the possibility in a randomized-controlled setting. The occurrence of neutralizing antibodies may call for the reconsideration of doses and dose regimens. Non-neutralizing antibodies may have a profound effect on PK and may therefore be just as important as neutralizing antibodies.

V. SPECIAL CONSIDERATIONS FOR MEDICAL DEVICES

A. Background

Medical devices used in the treatment of RA vary considerably in their therapeutic impact, ranging from primary therapy for a specific affected joint to products used as adjuncts to primary therapies such as the use of ultrasound or heat for symptomatic relief in conjunction with pharmaceutical therapy. Variability in therapeutic effects due to disease and response heterogeneity may be exaggerated with some devices because treatments are often targeted to a single or a few joints per patient.

Preclinical requirements for devices include assessment of data from a diverse range of tests, including, for example, tests that evaluate chemical composition, mechanical reliability or the electrical properties of a given device. Each product type has an identified battery of tests that are evaluated to ensure that the design of the specific product is well understood and can be expected to have the intended effect on patients as described in its labeling.

The design of clinical trials in RA for a device may raise some unique issues uncommon in pharmaceutical trials. The interaction between treated and untreated joints as it affects functional and quality of life assessments is just one example of these issues. Selection of an appropriate control often poses difficult questions when the active therapy under study is an implant and the available controls are more systemic in activity. Understanding how to design these trials, analyze the data, and assess the contribution from each therapy can be challenging.

B. Efficacy Considerations

1. The selection of a control group may be quite challenging with devices because masking for the patient and investigator is frequently impossible and a placebo may be infeasible. Because historical controls are often unsatisfactory and evaluating the equivalence of local therapy to a systemic active control poses problems, innovative trial designs are often needed. Some innovative options include randomization to early versus late device interventions, or rescue interventions following failure of pharmaceutical interventions. Although the use of a sham (placebo) device is the most desirable control for many products evaluated for RA, it may be inappropriate if the subject device is implanted. Additionally, the success of patient and/or physician masking with such shams may not always be complete. Patient masking is infeasible if the product requires a surgical or other invasive

procedure. However, in some cases it is possible to achieve evaluator masking for the effectiveness evaluation.

2. For devices intended for use as adjunctive therapy, approaches and analysis methods should be designed to account for differences in disease status and severity to minimize potential biases in outcome measures. One approach is to have a consistent primary therapy with a named pharmaceutical to avoid confounding the analyses. The use of additional, potentially confounding co-therapies (hot/cold treatments, splinting, physical therapy, and orthotics) also needs to be appropriately addressed during the study design phase.

3. The use of quality of life (QOL) assessments is very important where devices are intended for rehabilitative purposes. QOL benefits for the intervention should be judged with tools chosen for their validity, ease, and convenience of administration and the ability to address both improvement in QOL and product satisfaction. These assessments should be kept masked from the independent effectiveness evaluator to avoid assessment bias.

4. In the case of invasive devices necessitating in-hospital or in-office use, it is recommended that clinical assessments include convenience in use and pain or discomfort in administration. Early assessment of this aspect of therapy can provide critical information on the ultimate acceptability of the treatment and affect decisions as to whether to pursue a treatment modality.

C. Safety Considerations

1. Obtaining well-characterized, short-term adverse events rates, as an early assessment of safety as is common for large pharmaceutical trials is not always meaningful in medical device trials. Device trials are frequently much smaller in size and the reliability of early data may not be sufficient to establish a prospective pattern. When the device in question is an implant, the evaluation period must be fairly long to assesses the potential for late occurring adverse events including device failure and adverse reactions to device materials. In some cases, the full assessment of safety cannot be completed in the premarket phase and extends into mandatory postmarket surveillance.

2. When devices are used in conjunction with another medical or surgical procedure, the distinction between an adverse event that is clearly device related and one that is common to the procedure may be difficult to assess. The nature, timing, and degree of severity are some of the factors that may be useful in reaching a determination of cause. Since these determinations are based on clinical judgment, a real potential for bias exists in reaching conclusions about the treatment. Care must be exercised in designing the protocol to provide adequate detail in instruction to investigators to allow for consistent and unbiased decision making on their part.

3. Devices (e.g., those emitting radiation) that have the potential for intermittent or chronic use require assessment as to whether there is a lifetime exposure limit or a maximum frequency of exposure. Such issues are frequently addressed with animal data as well as with long-term clinical evaluations.

VI. SPECIAL CONSIDERATIONS FOR JUVENILE RHEUMATOID ARTHRITIS

A. Background

Juvenile rheumatoid arthritis (JRA) is a heterogeneous group of diseases that share the common feature of chronic, idiopathic synovitis, with onset prior to 16 years of age. These disorders have been divided into clinically distinct subsets based on the extent of joint involvement and extra-articular manifestations: pauci-articular, poly-articular, and systemic-onset JRA, as well as oligoarthritis associated with HLA-B27,[4] and they have been further subdivided based on clinical courses (Cassidy 1986). Immunogenetic subsets appear to correlate with these clinical course subsets and are also distinct from adult RA (Nepom 1991). Of these various entities, polyarticular JRA is similar in many aspects to adult RA, particularly in clinical signs and symptoms, synovitis, and similar efficacy responses to some existing pharmacotherapy (NSAIDs, methotrexate, and prednisone). As only 3 to 5 percent of all patients with rheumatoid arthritis develop illness onset during childhood, many investigational therapeutic agents in this small population will receive orphan drug status, according to 21 CFR part 316 — Orphan Drugs. The application of principles in the conduct of clinical trials for adult RA largely applies as well to JRA, and this section outlines only those areas of difference from adult RA. Sponsors are generally encouraged to develop as much information as possible on JRA patients for agents that will be approved for adult RA. As a minimum, dosing and safety data are strongly encouraged.

Conducting drug studies in children is generally necessary and consistent with the expectations of treatment regimens for this disease. Because pediatric subjects constitute a vulnerable population, conducting research involving minimal risk is important. The Committee on Drugs of the American Academy of Pediatrics has published guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations (AAP 1995a), and general considerations for the clinical evaluation of drugs in infants and children (AAP 1982), both of which should be consulted. Guidelines regarding informed consent and assent of pediatric patients from the Committee on Bioethics of the American Academy of Pediatrics should also be followed (AAP 1995b). Conducting clinical trials for patients with JRA and, particularly, assessing global disease activity and response to therapy should involve pediatric rheumatologists or adult rheumatologists who have extensive training in pediatric rheumatology and have demonstrated competence in caring for children with rheumatic diseases.

---

[4] The HLA-B27 subset is not addressed in this document.

As a general principle, children should not be subjected to an agent that has not been first tested for safety in adults. Testing may begin in children, however, when the anticipated benefits based on existing knowledge justify the anticipated risks. An agent developed specifically for use in JRA (e.g., a biological agent targeted against a specific pathogenic process that is unique to JRA and not present in adult RA) may need to be tested first in children, as exposure in adult RA patients or even normal adult volunteers may be unrevealing. If, however, the agent has potential for use in both adult RA and JRA, then, at minimum, PK-PD and initial phase 1 data (including MTD) for adults should be available prior to the start of testing in children. JRA trials of drugs that are expected to be similar in efficacy to existing drugs and that do not represent major therapeutic advances or alternative approaches to the basic mechanism of intervention can be delayed until there is extensive efficacy and safety data either from adults or in other pediatric populations.

The need for reliable inferences does not necessitate a placebo control, but randomization and controls should be employed. The choice of control is a function of what is known about the agent at the time and what other treatments are available to potential trial enrollees. If only an active control is used for an equivalence trial, convincing evidence of the efficacy of the active control should be provided, and the test proposed to establish equivalence should be specified. If there have been no prior adult studies, or if the agent under development has a novel mechanism of action or represents an entirely new class of drug, a randomized, double-blinded trial, using either a placebo or an active control group of (anticipated) similar efficacy is indicated. Open label extensions to obtain additional data about risk and persistence of benefit are very valuable. The use of active control (standard-of-care therapy) in the control arm, dose-response design (where control receives a lower dose(s) of the test agent), crossover, randomized withdrawal (enrichment design) or, if the agent has a short onset of effect, randomized placebo-phase trial designs are encouraged as possible alternatives to inactive placebo control in JRA studies (Temple 1994, Feldman 1995). As a general principle, protocol escape clauses are encouraged to permit children who are not responding well to experimental therapy to receive early conventional or alternative treatment. The sponsor should indicate how dropouts will be handled in the analysis, whether from the escape clause, or otherwise.

B. Applicability of Pediatric Regulation and Impact on Trial Design for JRA Studies

The *pediatric use* section in the labeling regulations (21 CFR 201.57(f)(9)) permits drug and biological products to be labeled for pediatric use if they have been demonstrated to be safe and effective for adult populations and the mechanism of action of the drug is sufficiently similar in children. The pediatric rule may be applied only to obtain labeling for the signs and symptoms of JRA; other claims, including prevention of structural damage, remission, and prevention of disability, should be evaluated in separate JRA efficacy studies. Although the regulation allows extrapolation of adult efficacy data, additional pediatric dosing and safety evaluations are usually needed.

In general, sponsors seeking approval for adult RA products appropriate for use in patients with JRA are strongly encouraged to obtain dosing and safety data in polyarticular course JRA for inclusion in the dosing and pediatric use sections of the label. Specimen collection for PK studies can be reduced significantly if available data indicate that the coefficients are similar in adults and children. Microsampling techniques should be employed for such studies (Hashimoto 1991). The extent of safety testing needed depends on the agent, its prior use, and any established safety in other pediatric populations. Toxicity grading scales should be adjusted for pediatric populations. Phase 4 studies for safety evaluation will be strongly encouraged when limited preapproval data are obtained. It is desirable that as much efficacy evidence as possible be gathered during the evaluation of pediatric dosing and safety.

For currently approved traditional (cyclooxygenase inhibitor) NSAIDs and corticosteroids, adequate efficacy information exists to support a labeled indication for all JRA and all JRA subsets. For methotrexate and sulfasalazine, adequate efficacy information exists for a labeled indication for JRA patients with a polyarticular course. For such agents, a labeling claim could be supported using only PK, PD, and safety data in JRA patients, although submission of additional JRA efficacy data is encouraged.

For new agents (not yet approved for adult RA) that are not from a new pharmacologic class, adult efficacy data can be used to support a signs and symptoms claim for polyarticular JRA if there is biological plausibility that the agent would have a similar effect in JRA. The applicability of the pediatric rule to support a labeled indication for polyarticular course JRA will be based on adult RA efficacy data considered on an individual basis for each agent. When evidence for biological plausibility does not exist, evidence should be submitted to support the application of the pediatric rule. (The Agency should be consulted in determining whether adequate biological plausibility exists to apply the pediatric rule.) Pediatric safety and dosing studies should be submitted.

For agents in a new class, efficacy studies should be performed in JRA to obtain an indication for use in JRA. The indication will reflect the JRA subsets included in the efficacy study. Sponsors who seek approval for all JRA should include all JRA subsets in an efficacy study. The data could support a claim for JRA (subsets not specified) provided that the data do not suggest that the agent is ineffective in any one subset. The label should reflect that efficacy was demonstrated and that the agent is approved for JRA (subsets specified depending on which were included in the efficacy study).

C. Outcome Variables and Claims

It is possible for sponsors to seek approval for all JRA subsets or to seek approval for individual subsets. In the former case, the label should note the number of patients from each subset enrolled in trials and the character of each subset response. Except as noted above in the application of the pediatric rule, all claims should be supported by an efficacy demonstration in the intended subset(s).

1. Reduction in the Signs and Symptoms of JRA

All JRA trials should evaluate improvement based on a validated endpoint for improvement. Currently, the one validated approach is the definition of improvement established by the JRA core set: three of six (MD global, parent/patient global, number of active joints, number of joints with limited range of motion, functional ability, and ESR) improved by at least 30 percent and no more than one of six worsening by more than 30 percent (Giannini 1997). Because the JRA definition of improvement was validated using a trial of methotrexate, which primarily included polyarticular JRA patients, protocol individualization may necessitate a refinement in the responder test for other patient subsets. For example, for pauci-articular JRA, with one knee involved and a normal ESR, use of joint and functional assessments specific to the involved joints and evaluation of uveitis as coprimary endpoints may also be valuable (Lindsley 1996). For patients with systemic onset JRA, additional assessment of fever, extra-articular manifestations, and thrombocytosis/leucocytosis may be useful coprimary endpoints (Silverman 1994). Outcome variables need to be appropriate and consistent with the type of agent under investigation. Investigators should specify, before the trial is initiated, how much change is considered clinically important for each outcome variable.

Trials should generally last at least six months, except when six-month efficacy data exist in adult RA and there are no reasons to expect loss of efficacy over time. Under these circumstances, trial durations may be three months' blinded/randomized, but six-months' open safety data should be obtained. As with adult RA, a three-month trial duration is suggested for NSAIDs.

2. Major Clinical Response

Similar to adult RA, major clinical response is a claim intended to connote that the agent provides substantial clinical benefit, including in patients who are unable to completely respond to the treatment or remit from the disease. At present, this claim is only theoretical, as clinical JRA trial databases adequate for defining major clinical response do not exist.

3. Complete Clinical Response

The claim of complete clinical response reflects achievement on drug of six consecutive months of morning stiffness of less than 15 minutes duration, no active synovitis (pain, redness, tenderness to palpation, swelling, stable or decreasing limitation of motion), no extra articular features (including fever, serositis, adenopathy, hepatosplenomegaly, rash, uveitis), and normal laboratory parameters (including ESR, platelets, WBC) and, where applicable, no ongoing structural damage while continuing on therapy. Trials should be at least one year in duration. Residual damage from prior disease, including extra articular manifestations, is

acceptable in meeting criteria for complete clinical response. Because spontaneous complete clinical response rates may be relatively high in JRA, these studies should be controlled.

4. Remission

Remission is characterized exactly as above, but while off all antirheumatic drugs. Remission is not intended to imply cure.

5. Prevention of Disability

This claim is proposed to reflect durable improvement in physical function and disability in studies of one to two years' duration with demonstrated improvement in signs and symptoms over the same period. Instruments currently validated for use in JRA include the Childhood Health Assessment Questionnaire (CHAQ), the Juvenile Arthritis Self-Report Index (JASI), and the Juvenile Arthritis Functional Assessment Report (JAFAR). HR-QOL should also be measured and demonstrated not to worsen over the trial duration. Endpoints should be tailored to subtypes enrolled in trials (e.g., to assess knee function in pauci-articular JRA patients in whom knee arthritis may be the primary arthritic manifestation). Instruments should be developmentally validated for the age ranges studied in a trial (Murray 1995).

6. Prevention of Structural Damage

Similar to adult RA, this claim would reflect trials of one year or more with concomitant success in signs and symptoms. Currently, only sparse data exist regarding the usefulness of only one radiographic measure in JRA: the carpal-metacarpal distance in those patients with wrist arthritis. Other clinically promising settings include the evaluation of erosive disease in systemics with polyarthritis, hip assessment in systemics, and knee assessments in pauci-articular JRA.

D. Trial Design Issues

Recommendations for efficacy studies are based on the nature of the agent under development. The principles outlined for adult RA are generally applicable. Patients enrolled in these trials may be of any onset or disease course subset. Separate trials for each JRA subset are recommended if the agent is predicted to have a target mechanism of action that will not be applicable and equally efficacious in all JRA subsets. Alternatively, a single, sufficiently large trial with enrollment appropriately stratified provides for useful conclusions to be reached about efficacy and safety for each subset. Relevant covariates include disease course type, disease duration, and nonresponse to prior methotrexate treatment. Given that JRA is an orphan disease, there is often some flexibility in trial design, but this should be discussed on a case-by-case basis.

At this time, JRA patients are usually ineligible for entry into efficacy trials unless they have failed to respond adequately to at least one standard *second line agent* (such as methotrexate at a dose of at least 10 mg/m^2 body surface area per week). There may be exceptions to this if, for example, there is evidence that greater efficacy could be obtained by using the agent very early in the disease course, evidence that delayed use in sicker patients potentially carries greater risk of toxicity, or evidence that the agent has a favorable safety and efficacy profile in a comparable population studied to date and that the agent's actions are potentially readily reversible. Pauci-articular JRA patients are particularly encouraged for inclusion in trials with agents targeting the treatment of uveitis or agents that will replace existing therapy with an improved safety profile, less frequent blood monitoring, and/or superior efficacy.

Some JRA patients exposed to new agents should be evaluated periodically for an extended period. Effects on skeletal growth, development, behavior, sexual maturation, reproductive capacity, and secondary malignancy should be included in such monitoring. Registries or cohort follow-up studies may be useful in providing long-term safety information.

E. Concurrent Antirheumatic Agent Administration

The principles of use of concurrent antirheumatic therapy in JRA trials are similar to those outlined for adult RA: limiting their discretionary use as much as reasonably possible so that interpretation of efficacy and safety data is not compromised. However, limitations on concurrent medication cannot prohibit ethically justified treatments, nor should the protocol be made so unattractive to parents, physicians, and patients that enrollment is threatened. If background treatment is necessary, early tolerance studies, to ensure safety of co-administration, should precede any large trials.

If patients receive concurrent slow-acting or prednisone therapy, the dose should be stable prior to study entry and should preferably remain so throughout the trial. Concurrent medications are usually important prognostically and so may need stratification. If possible, intra-articular steroid injections should be disallowed for a minimum of one month prior to beginning experimental therapy; otherwise, that joint should be discounted in assessing therapeutic effects.

F. Multicenter Trials and Center Effects

Although JRA is the most common rheumatic disease of childhood, its prevalence is low compared to adult RA. Thus, trials of JRA that require large numbers of patients will likely be multicenter trials. Multicenter trials should employ a standardized protocol and data collection forms among all centers. Pretrial meetings of all investigators and other involved personnel are strongly encouraged to ensure uniformity in protocol interpretation, patient evaluation, and data recording. Studies have shown that, within a cooperative group, a center's performance is a function of the number of patients enrolled at the center (Sylvester 1981). Thus, studies that use fewer centers with greater numbers

of patients at each center are preferable to those that use large numbers of centers with fewer patients. Effort should be made to enroll at least 10 to 12 patients at each center to provide for greater quality assurance. In all multicenter trials, center effects should be examined. A therapy should show effect in more than one center. When stringent entrance criteria restrict the number of patients eligible for study, many centers may be unable to enroll even 10 patients. In such situations, randomization blocked within individual centers, rather than across all centers, may help to reduce the potential impact of center effects.

REFERENCES

American Academy of Pediatrics (AAP), Committee on Drugs, 1982, *General Considerations for the Clinical Evaluation of Drugs in Infants and Children*, Report to the U.S. Food and Drug Administration, available through the Freedom of Information Office of the FDA, FDA Contract Nos. 223-79-3003 & 223-82-3009.

_____, Committee on Drugs, 1995a, "Guidelines for the Ethical Conduct of Studies to Evaluate Drugs in Pediatric Populations," *Pediatrics*, 95: 286-294.

_____, Committee on Bioethics, 1995b, "Informed Consent, Parental Permission, and Assent in Pediatric Practice," *Pediatrics,* 95: 314-317.

Cassidy, J.T., J.E. Levinson, J.C. Bass, et al., 1986, "A Study of Classification Criteria for a Diagnosis of Juvenile Rheumatoid Arthritis," *Arthritis and Rheumatism*, 29: 274-281.

Feldman, B.M. and J.P., Szalai, 1995, "Towards a More Ethical Clinical Trial Design: The Randomized Placebo Phase Design (RPPD)," abstract, *Arthritis Rheum*, 38 (supplement): s177.

Felson, D.T., et al., 1993, "The American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis Clinical Trials," *Arthritis and Rheumatism,* 36(6): 729-740.

Felson, D.T., et al., 1995, "American College of Rheumatology Preliminary Definition of Improvement in Rheumatoid Arthritis," *Arthritis and Rheumatism*, 38(40): 1-9.

Food and Drug Administration, 1978, *General Considerations for the Clinical Evaluation of Drugs*, FDA 77-3040.

_____, 1988, *Guidelines for the Clinical Evaluation of Anti-Inflammatory and Antirheumatic Drugs (Adults and Children).* April 1988.

_____, 1997, *Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro,* April 1987.

_____, 1993, "Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs," *Federal Register,* July 22, 1993, 58 (139): 39406-39416.

Fries, J.F., P.W. Spitz, and D.Y. Young, 1982, "The Dimensions of Health Outcomes: the Health Assessment Questionnaire, Disability, and Pain Scales," *Journal of Rheumatology,* 9: 789-793.

Giannini, E.H., N. Ruperto, A. Ravelli, et al., 1997, "Preliminary Definition of Improvement in Juvenile Arthritis," *Arthritis and Rheumatism,* 40: 1202-1209.

Hare, L., L. Wagner-Weider, A. Poznanski, et al., 1993, "Effects of Methotrexate on Radiologic Progression in Juvenile Rheumatoid Arthritis," *Arthritis and Rheumatism*, 36: 1370-1374.

Hashimoto, Y., and L.B. Sheiner, 1991, "Designs for Population Pharmacodynamics: Value of Pharmacokinetic Data and Population Analysis," *Journal of Pharmacokinetics and Biopharmaceuticals*, June 1991, 19(3): 333-53, NLM CIT. ID: 91341632.

ICH, "S5A Guideline on Detection of Toxicity to Reproduction for Medicinal Products," *Federal Register*, Thursday, September 22, 1994, 59 (183): 48746-48752.

_____,"S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals," *Federal Register*, April 4, 1997, 62 (65): 16437-16442.

_____,"EIA Extent of Population Exposure Required to Assess Clinical Safety for Drugs Intended for Long-Term Treatment of Non-Life-Threatening Conditions," *Federal Register*, March 1, 1995, 60 (40): 11269-11271.

Larsen, A., K. Dale, and M. Eek, 1977, "Radiographic Evaluation of Rheumatoid Arthritis and Related Conditions by Standard Reference Films," *Acta Radiol. Diag.* 18: 481-491.

Lindsley, C., 1996, "Juvenile Rheumatoid Arthritis Workshop," Proceedings, pp. 34-43.

Little, R., and D. Rubin, 1987, Statistical Analysis with Missing Data, John Wiley & Sons.

Meenan, R.F., P.M. Gertman, J.H. Mason, et al., 1982, "The Arthritis Impact Measurement Scales: Further Investigations of Health Status Measure," *Arthritis and Rheumatology,* 25: 1048-1053.

Murray, K.J., and M.H. Passo, 1995, "Functional Measures in Children with Rheumatic Diseases," *Pediatric Clinics of North America*, 42: 1127-1154.

Nepom, B., 1991, "The Immunogenetics of Juvenile Rheumatoid Arthritis," *Rheumatic Disease Clinics of North America*, 17: 825-842.

Paulus, H.E., M.J. Egger, J.R. Ward, et al., 1990, "Analysis of Improvement in Individual Rheumatoid Arthritis Patients Treated with Disease-Modifying Antirheumatic Drugs Based on Findings in Patients Treated with Placebo," *Arthritis and Rheumatism,* 33: 477-484.

Pinals, R.S., et al., 1981, "Preliminary Criteria for Clinical Remission in Rheumatoid Arthritis," *Arthritis and Rheumatism*, 24 (10): 1308-1315.

Sharp, J.T., D.Y. Young, G.B. Bluhm, et al., 1985, "How Many Joints in the Hands and Wrists Should Be Included in a Score of Radiologic Abnormalities Used to Assess Rheumatoid Arthritis?" *Arthritis and Rheumatism,* 28: 1326-1335.

Silverman, E., G.D. Cawkwell, et al., 1994, "IVIG in the Treatment of Systemic Onset JRA," *The Journal of Rheumatology,* 21(12): 2353-2358.

Smolen, J.S., F.C. Breedveld, G. Eberl, et al., 1995, "Validity and Reliability of the Twenty-Eight-Joint Count for the Assessment of Rheumatoid Arthritis Activity," *Arthritis and Rheumatism,* 38: 38-43.

Sylvester, R.J., H.M. Pinedo, M. De Pauw, et al., 1981, "Quality of Institutional Participation in Multi Center Clinical Trials," *New England Journal of Medicine*, 305: 852-855.

Temple, R.J., 1994, "Special Study Designs: Early Escape, Enrichment, Studies in Non-Responders." *Comm. Statistic. Theory Meth.* vol 23(2): 499-531.

APPENDIX A: COMPARATIVE TRIAL RESPONSE RATES [5]

Three cooperative systematic studies of rheumatic diseases (CSSRD) trials:
(1) Methotrexate vs. Placebo
(2) Gold, Auranofin vs. Placebo
(3) D-Penicillamine high, low vs. Placebo

Response Rates at End of Trial Based on Different Definitions of Improvement

Definition of Improvement	PLACEBO n=199	(Auranofin) (Low-dose D-Penicillamine) n=18	High-dose D-Penicillamine Gold, Methotrexate n=155
ACR≥ 20%	10 (8.4%)	30 (25.4%)	64 (40.3%)
ACR≥ 30%	.5 (4.2%)	14 (12.0%)	46 (29.7%)
ACR≥ 40%	2 (1.7%)	7 (3.4%)	18 (11.6%)
ACR≥ 50%	0 (0%)	4 (3.4%)	14 (9.0%)
ACR≥ 60%	0 (0%)	3 (2.5%)	4 (2.6%)
ACR≥ 70%	0 (0%)	0 (0%)	1 (0.6%)

[5] CSSRD database and Tugwell et al., "Combination Therapy with Cyclosporine and Methotrexate in Severe Rheumatoid Arthritis," *N Engl J Med* 333;137-141, 1995.

APPENDIX A (cont.)

COMPARATIVE MULTICENTER TRIAL OF AURANOFIN/METHOTREXATE (END OF TRIAL)

Response Rates at End of Trial Based on Different Definitions of Improvement

Definition of Improvement	Auranofin (N=118)	Methotrexate (n=119)
ACR≥20%	34 (28.8%)	77 (64.7%)
ACR≥30%	30 (25.4%)	65 (54.6%)
ACR≥40%	22 (18.6%)	51 (42.9%)
ACR≥50%	21 (17.8%)	42 (35.3%)
ACR≥60%	9 (7.6%)	22 (18.5%)
ACR≥70%	7 (5.9%)	11 (9.2%)

APPENDIX A (cont.)

COMPARATIVE TRIAL OF CYCLOSPORINE-A METHOTREXATE VS. METHOTREXATE ALONE

Response Rates at End of Trial Based on Different Definitions of Improvement

Percent Increase by ACR Criteria	Patients Satisfying Criteria	
	Methotrexate + Cyclosporine-A (n=71)	Methotrexate + Placebo (n=74)
0%	81.7	50.0
10%	49.3	16.2
20%	45.0	12.2
30%	33.8	8.1
40%	22.5	2.7
50%	22.5	2.7
60%	5.6	2.7
70%	1.4	0.0
80%	0.0	0.0
90%	0.0	0.0