CENTER FOR DRUG EVALUATION AND RESEARCH

Guidance for Industry

The FDA published Good Guidance Practices in February 1997. This guidance was developed and issued prior to that date.

Additional copies are available from:
Office of Training and Communications
Division of Communications Management
Drug Information Branch, HFD-210
5600 Fishers Lane
Rockville, MD 20857

(Tel) 301-827-4573
(Internet) http://www.fda.gov/cder/guidance/index.htm

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, FOOD AND DRUG ADMINISTRATION

GUIDELINES FOR THE CLINICAL EVALUATION OF

ANTI-INFLAMMATORY AND ANTIRHEUMATIC DRUGS
(ADULTS AND CHILDREN)

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, DC 20402

GUIDELINES FOR THE CLINICAL EVALUATION OF

ANTI-INFLAMMATORY AND ANTIRHEUMATIC DRUGS
(ADULTS AND CHILDREN)

September 1977
Revised April 1988

Comments on the contents of this publication are invited and should be addressed to the following office, and identified with the docket number:

Dockets Management Branch (HFA-305)
Docket Number 88D-0050
Rm. 4-62
5600 Fishers Lane
Rockville, Maryland 20857

For further information regarding the guidelines please contact:

Group Leader for Anti-inflammatory Drugs
Division of Oncology and Radiopharmaceutical Drug Products (HFN-150)
Office of Drug Evaluation
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, Maryland 20857

TABLE OF CONTENT

ABSTRACT

The Food and Drug Administration, with the assistance of its scientific Advisory Committees and other outside consultants, and the American Academy of Pediatrics' Committee on Drugs, has developed guidelines for the clinical evaluation of the new drugs. These guidelines present acceptable current approaches to the study of investigational drugs in man, and pertain to Phases I through III of the investigation. They represent generally accepted principles for arriving at valid conclusions concerning safety and effectiveness of new drugs, as well as the views of outstanding experts concerning appropriate methods of study of specific classes of drugs.

The FDA welcomes comments on the guidelines, and expects to keep them current by review and update.

FOREWORD

The purpose of these guidelines is to present acceptable current approaches to the study of investigational drugs in man. These guidelines contain both generalities and specifics and were developed from experience with available drugs. It is anticipated that with the passage of time these guidelines will require revision. An introduction to this second edition appears on page 1.

These guidelines are not to be interpreted as mandatory requirements by FDA to allow continuation of clinical trials with investigational drugs or to obtain approval of a new drug for marketing. These guidelines, in part, contain recommendations for clinical studies that are recognized as desirable approaches to be used in arriving at conclusions concerning safety and effectiveness of new drugs; and in the other part they consist of the views of outstanding experts in the field as to what constitutes appropriate methods of study of specific classes of drugs. In some cases other methods may be equally applicable or newer methods may be preferable, and for certain entirely new entities it is possible that the guidelines may be only minimally applicable.

Under FDA regulations [21 CFR 10.90(b)], clinical guidelines constitute advisory opinions on an acceptable approach to meeting regulatory requirements, and research begun in good faith under such guidelines will be accepted by the Agency for review purposes unless the guideline (or the relevant portion of it) has been formally rescinded. This does not imply that results obtained in studies conducted under these guidelines will necessarily result in the approval of an application or that the studies suggested will produce the total clinical information required for approval of a particular drug.

Many of the clinical guidelines have been developed largely, or entirely, by FDA's Advisory Committees and consultants. Others were originally developed by intramural committees and consultants of FDA, the Pharmaceutical Manufacturers Association, and members of various professional organizations; in these cases, the guidelines were reviewed and revised, as appropriate, by FDA's staff and Advisory Committees.

The general guidelines for the evaluation of drugs in infants and children and most of those for study of various drug classes in children were developed by the Committee on Drugs of the AAP (American Academy of Pediatrics). Some of the pediatric guidelines for specific classes were written by FDA's Advisory Committees. There was cross-review and comment on the pediatric guidelines by both the Committee on Drugs of the American Academy of Pediatrics and the Advisory Committees of the FDA.

FDA's Office of Drug Evaluation wishes to thank the many individuals who spent so much time and effort to the develop these guidelines.

Robert Temple, M.D.
Director, Office of Drug Evaluation I

GUIDELINES FOR THE CLINICAL EVALUATION OF

ANTI-INFLAMMATORY AND ANTIRHEUMATIC DRUGS - (ADULTS AND CHILDREN)

INTRODUCTION

This is the second edition of guidelines for the clinical evaluation of anti-inflammatory drugs. There is a change in the title from "Guidelines for the Clinical Evaluation of Anti-inflammatory drugs (adults and children)" to "Guidelines for the Clinical Evaluation of Anti-inflammatory and Antirheumatic Drugs (adults and children)". The first edition, published in 1977, contained only guidelines for evaluating nonsteroidal anti-inflammatory drugs (NSAIDs) while this edition also contains guidelines for evaluating Disease Modifying Antirheumatic Drugs (DMARDs).

Although revision of guidelines occurs only at several year intervals, new information accumulates continuously. To keep abreast of, and anticipate changes in the guidelines, it would be useful to follow the activities of the Arthritis Advisory Committee and its subcommittees. Meetings are announced and agenda presented in the Federal Register in the month before the meetings take place. Moreover, while these guidelines are relatively detailed, they cannot replace careful consideration of individual protocols and overall plans with outside expert consultants and FDA staff.

All the NSAIDs studied to date have had analgesic properties as well as anti-inflammatory and anti-pyretic effects. Several members of the class have been approved solely for analgesic indications. Persons interested in the clinical evaluation of NSAIDs for analgesic indications, including dysmenorrhea, should consult the Guidelines for the Clinical Evaluation of Analgesic Drugs. Guidelines have been proposed for developing and submitting NDAs for fixed dose combinations of NSAIDs with codeine. They are included as Appendix 7 of these guidelines at the end of the section on NSAIDs.

RHEUMATOID ARTHRITIS -- ADULT SECTION

I. INTRODUCTION

A. Relevant Guidelines

"General Considerations for the Clinical Evaluation of Drugs" is an important companion document and should be reviewed prior to reading these guidelines. It contains suggestions that are applicable to investigational drug studies for most classes of drugs and enables elimination of repetitious material in each of the specific guidelines. In addition, the "Guideline for the Format and Content of the Clinical and Statistical Sections of an Application" can be used to anticipate what information will need to be collected and presented.

B. General Comments

There are four important target areas available for pharmacological attack in the clinical management of inflammatory states. These areas are:

1. The etiological factor (primary cause);

2. Mediators of the initial tissue injury released by, or produced in response to, the etiological factor (these include fundamental mechanisms for initiation and perpetuation of the secondary inflammatory response);

3. The nonspecific inflammatory response evoked by this tissue injury; and

4. The processes attempting to restore normal function.

The methods of investigation, realistic goals, and inherent toxicity of therapy may differ in many aspects for each of these four areas.

The guidelines below are appropriate for a prompt-acting agent in an established inflammatory process (No. 3 above), such as active rheumatoid arthritis. Aspirin is an example of such an agent. Prompt-acting agents may effectively reduce the signs and activity

within the first few days of administration and may only be effective when blood levels are sustained; their withdrawal is soon associated with recurrence of signs and symptoms, especially the pain and general subjective discomfort associated with rheumatoid arthritis. Gold salts and other slow-acting agents are not prompt-acting agents. Such agents are directed at the initiation of the inflammatory process, and are sometimes referred to as disease-modifying antirheumatic drugs (DMARDs). If this activity is being considered, refer to the appropriate section of the guidelines.

Major emphasis is placed on rheumatoid arthritis because:

1. It is a good example of musculoskeletal inflammation that is not self-limited, is relatively unresponsive to placebo, and less variable from day to day than are most of the other disease processes that might be considered.

2. It is a prevalent condition and under study by many physicians and clinics.

3. Historically, anti-inflammatory drugs have been shown to have substantial effects in rheumatoid arthritis.

4. The American Rheumatism Association, through its committees, has defined and standardized the diagnosis of rheumatoid arthritis and various assessments of disease activity. These have been widely and successfully used, so that both clinical investigators and FDA feel confident that they can evaluate new anti-inflammatory drugs using patients with rheumatoid arthritis.

II. PRECLINICAL STUDIES

Appropriate preclinical studies should be conducted to assess the dose at which toxicity appears, and the target organs involved, as well as the dose-response results in various animal models used to screen drugs for pharmacological activity, before initiating clinical studies (see preclinical guidelines on human drugs).

CLINICAL STUDIES

A. General Considerations

There are a number of general principles that relate to clinical studies of anti-inflammatory drugs that apply to the other conditions discussed later in these guidelines, as well as to rheumatoid arthritis.

1. Clear definition of study objective and critical variables of effectiveness.

   Clinical trials of anti-inflammatory drugs often compare two or more treatment entities, using a number of measures of disease activities, evaluated at different time points during the trial. These multiple comparisons are difficult to analyze when the study objective is too broad and imprecise, such as to compare the experimental drug to the control drug for the treatment of the symptoms of rheumatoid arthritis. The study objective should be sharply defined and precisely stated, especially when the objective is to test for a specific therapeutic effect of a drug as suggested by <u>in vitro</u> animal and pilot human studies.

   The primary efficacy variables, i.e., those measurements that are considered most critical in assessing the effectiveness of the drug, should be those disease manifestations that most consistently and accurately reflect disease activity and the changes caused by the disease in the patient's functional capability. They should be clearly related to the proposed indication for the drug. The protocol should include a clear description of what is to be measured, as well as how and when measurements are to be made, and which variables are to be followed as the primary efficacy parameters.

   Experience has shown that some efficacy outcome variables are more sensitive and reliable than others (e.g., physician's and patient's global evaluation of disease activity, swollen and tender joint counts in rheumatoid arthritis, and patient's assessment of pain) and these variables should always be assessed, even if additional variables are also of interest. Because statistical testing of multiple measures of disease activity at multiple times can lead to misleading chance findings of apparent statistical significance, three to four efficacy variables should be selected as primary efficacy variables, to be analyzed at two to three preselected time points. Other measures of disease activity (see Appendix 3) and measurements at other time points should be followed as secondary or supportive measures of effectiveness.

2. Clear definition of use and evaluation of concomitant medications.

   Concomitant use of arthritic medication and analgesics in a trial of anti-arthritic drug may confound the drug effects and introduce bias into comparisons. Therefore, whenever possible, their use should be prohibited in the study. If this is not feasible, their use should be carefully specified in the protocol and arrangements made to supervise and monitor medications. Their consumption should be accounted for in the efficacy and safety evaluation. As an alternative approach, the study can be designed and conducted to incorporate the need for certain specified concomitant medication as one of the primary efficacy measurements.

3. Provisions for handling of dropouts, exclusions, and missing data.

A common problem in studying anti-inflammatory drug trials is how to deal with the increasing number of dropouts as the study progresses, and how to assess the impact of differences between treatment groups in the number of dropouts. Even in a well-conducted study that does not lose many patients for administrative reasons, ineligibility, protocol violations, and noncompliance, there are many dropouts in most trials, the main causes being lack of efficacy and adverse drug reactions. While these may be unavoidable, one cannot assume that dropouts, with censoring of subsequent observations and missing data, occur at random and do not affect the outcome. Statistical inference, based on the available data alone and ignoring the missing ones, may be biased. To cite a recurring example, placebo-controlled trials in rheumatoid arthritis typically have many dropouts in the placebo group due to lack of effectiveness. If only "one-month-completers" are compared, the analysis is highly biased against the active drug, as the patients who did poorly are excluded.

Some preventive tactics can be incorporated in the study design to minimize unnecessary loss of patients:

(a) Screening patients initially to eliminate patients with little or no commitment to participate in the study.

(b) In studies with a washout period, assigning patients to treatment groups only after their baseline evaluation is completed and their eligibility for the study is confirmed.

(c) In long-term trials with active control treatments, choosing a dosage for the control drug to minimize dropouts.

(d) Training investigators so that all patients entered are eligible, concomitant therapy is detected and accounted for, and the protocol is not violated.

The study protocol should specify the procedure for dealing with dropouts and exclusions in the statistical analysis. Studies, especially those with placebo groups, may be designed to use the dropout rate itself as one of the primary efficacy variables to assess the effect(s) of therapy. The time of withdrawal and the reason(s) for dropping out are critical aspects to consider in utilizing dropouts for determining the therapeutic value of the treatment, in terms of both benefits and adverse effects.

In addition to thorough planning for dropouts, presentation of study results should include a thoughtful discussion of how dropouts, exclusions, and methods for handling missing data could have affected the outcome(s) of the study. The discussion should include: (a) a demonstration that the results are consistent and do not depend on the method used to incorporate dropouts and missing data, i.e., that significant treatment effects are found using either an endpoint analysis (last available patient evaluation) or an analysis of data for only patients who completed the entire treatment period; (b) a comparison of treatment groups with respect to the number of dropouts due to lack of efficacy, adverse reactions, loss to follow-up, etc.; (c) an assessment of results including and excluding data for protocol violators and other patients that cannot be evaluated after enrollment; and (d) consistencies in the result of analyses at multiple time points.

4. Evaluation of the effect of differences in baseline disease activity.

   Patients with different degrees of severity of disease, or certain characteristics, may respond to the drug differently. When the patient population at baseline is not homogeneous with respect to a variable of concern, it is necessary, as a minimum, to examine the effect on the analysis of this variability and, in some circumstances, to formally include the baseline severity in the evaluation.

   The efficacy comparison can be adjusted for the baseline effect through covariate analysis. The covariate analysis should be conducted with a method or model that is suitable to the problem. Assumptions of the model should be formally examined.

5. Evaluation of the response of the active control group in positively controlled studies.

   In general, the comparison of a test drug and an active control is most meaningful if both are administered at their optimal dose levels within the context of intended therapeutic benefits and anticipated adverse drug effects.

   There are situations where use of a placebo control is undesirable, e.g., long-term evaluation of a new drug, or where comparative data are sought for labeling purposes to include information relating the new drug to standard treatment. In such cases, an active drug may be selected to serve as a control. The control treatment selected should be one whose clinical use for the disease under study is well documented and whose therapeutic effects are well understood, so that its effectiveness in the trial as a control can be assured in the

absence of a concurrent placebo control. This is especially critical when the objective of the study is to show equivalence (or lack of difference) between the test drug and the active control. In that case, the planned analysis should include an evaluation of the control group response. If the control group does not reach the expected level of response, the study may be regarded as a methodological failure.

Assuming the active control appears to have had an effect, there needs to be an analysis of the ability of the study to have detected a difference between the treatments. One approach to do this is to calculate the lower 95%-confidence limit of the ratio [test drug improvement from baseline divided by control drug improvement from baseline (=q)] for the primary efficacy variables. In general, studies of adequate size including drugs with similar effects have given q-values above .60 for studies in patients with rheumatoid arthritis. In other words, studies should provide 95% confidence that the test drug is at least 60% as effective as the control.

When the objective of an active control study is to demonstrate the superiority of the test drug to the control, the analytical procedures are relatively straightforward and are similar to those in placebo-controlled trials. All characteristics of the response distributions should be considered, however, not just those that favor the test drug. Similarities as well as dissimilarities between the two drugs need to be examined.

6. Selection of sample size.

   The study protocol should discuss the number of patients required to yield a meaningful result. When the study is designed to show a difference between treatments, an inadequate sample size will result in an excess likelihood of a Type II error, an outcome which is obvious from the trial result. When the study is intended to show no difference, great care must be taken to avoid excessive Type II error, i.e., a failure to statistically detect an important difference between the test drug and the active drug.

7. Justification for any data transformation.

   The observed data of each efficacy variable may not satisfy all the assumptions of the statistical method intended to be used for analysis. Sometimes a transformation can be applied to normalize the data, to homogenize the variances in various cells of the classification of data, etc. Before transforming the data for analysis, one must discuss the characteristics of the transformation and how to translate the inference made in the transformed probability space to the original framework. If the meaning of the new parameters, new statistical hypotheses, as well as the

statistical results based on the transformed data cannot be informatively and accurately interpreted, then it is essential also to display and interpret the results of the original untransformed response data.

8. Analysis of adverse reaction and laboratory data.

Statistical approaches used to evaluate adverse reaction data and laboratory values are usually different from those used to evaluate the efficacy data. The purpose of safety evaluations is rarely to test a specific hypothesis; instead, it is to examine the pattern of effects and to detect unusual events. Exploratory techniques such as life-table estimates (now called Time-related Cumulative Occurrences = TCO) of adverse reaction rates over time, scatter-plots of laboratory values (baseline versus on-therapy values), or general regression techniques may be helpful. This type of analysis will provide useful labeling information and may also generate new hypotheses to be examined in future studies.

Importantly, the exploratory analysis should address the extent to which adverse drug reactions and abnormal lab values depend on the duration of drug use, dose level, the patient's underlying condition, or possible drug interactions. Calculation of the incidence of adverse reactions should be based on a denominator that reflects the period of drug exposure for the population at risk. The cumulative incidence or the hazard rate (instantaneous probability) for adverse reactions can better represent the temporal pattern of drug effects than does the prevalence rate.

If possible, an attempt should be made to characterize the patient population susceptible to, or most susceptible to, any adverse drug effects. Some extraneous factors can complicate the safety data, such as variations in soliciting and reporting adverse reactions among the investigators, and differences in the definition of normal ranges for lab values among different laboratories. 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.

B. Special Studies and Considerations

Apart from studies of effectiveness and safety in the treated population of interest, NSAIDs pose a number of special problems that must be addressed either in the course of clinical pharmacology evaluation or during clinical studies.

Clinical Pharmacology

1. As a class, NSAIDs have potentially serious toxicity that is

dose-related, and individual agents have often had unique toxicities. In addition, the drugs are given to a population that includes the elderly, who have other illnesses, who frequently receive other drugs, and who often have some degree of impaired renal excretion or hepatic metabolism. It is then particularly important to evaluate, early in drug development, the metabolism and excretion patterns of the drug, and to study the effects of renal or hepatic impairment, age, and other therapy on the pharmacokinetics of the drug. These drugs are particularly good candidates for population kinetic studies.

2. Studies of observed effects on specific organs and systems, including long-term ophthalmologic examination and measures of potentials for gastrointestinal bleeding, both in short-term studies and long-term trials should be conducted. Other effects should be evaluated, when indicated, by preclinical tests, or by results from earlier trials.

3. The protein binding of the new drug and of its metabolism should be assessed.

4. As NSAIDs are intended for the long-term treatment of a large population, patient exposure during investigation of the drug should be substantial and should include considerable long-term exposure, i.e., 200 to 400 patients for 1 year, 100 to 200 patients for 2 years.

5. Because NSAIDs, as a class, have relatively low therapeutic indices, it is particularly important to characterize the dose-response relationship for effectiveness and for specific toxicities.

IV. PHASE I STUDIES

Aside from studies of pharmacokinetics and drug metabolism, and other clinical pharmacology studies in human subjects, Phase I is used to determine how single doses are tolerated, and to assess the tolerance of multiple doses for relatively short periods (4 to 6 weeks). A preliminary assessment of effectiveness and an approximation of the therapeutic dose can be gained from open studies in patients with very active disease (never treated or allowed to flare) by physicians very experienced with anti-inflammatory and antirheumatic drugs. Upward titration should be fairly rapid (e.g., weekly) in such patients. Experience in the first patients treated may lead to increases in starting doses in subsequent patients.

V. PHASE II STUDIES

Phase II is intended to provide unequivocal evidence of effectiveness from well-controlled studies, and to begin to explore the dose-response curve for the drug. The earliest studies should be placebo-controlled studies in patients with active rheumatoid arthritis, as defined in Appendices 1 and 2,

and may be short-term, 4 to 6 weeks. A dose reasonably well tolerated and assessed as probably effective in Phase I studies should be selected, but it may be necessary to carry out further studies at lower or higher doses, depending on early results. It may be most efficient to design early controlled, parallel dose-response studies, comparing several doses of the test drug with placebo. As a practical matter, a number of studies each, with only two dosage groups per study, comparing three- to fourfold differences in dosage, may be more successful than a single study with the same total number of patients using a larger number of more closely spaced doses. With many dosing groups, each investigator in a multicenter study will have only a few patients in each treatment group. The objective at this stage is to begin to define both the dose-effectiveness relationship and the dose-tolerance relationship, to the extent this can be done in short-term studies. Refinements of dosage can and should extend into Phase III.

Comparisons with other agents can begin in late Phase II, after a reasonable idea of dose-response is obtained. These studies should also be double-blind comparisons and should compare either the optimal dose of the test drug with a defined optimal dose of control agent or, even better, a number of doses of each agent, with all doses in the range known to be effective. The comparison with aspirin, for example, should usually use a dose of at least 3.6 g/day. These studies should be 6 to 12 weeks in duration.

In all studies, single entity non-anti-inflammatory analgesics (no combinations) may be used, but dose and frequency should be recorded carefully.

Patients should be monitored at least biweekly during the first 6 weeks of Phase II studies for tolerance and effectiveness, including the effectiveness assessment set forth in Appendix 3, physical examination and recording of adverse events, and laboratory evaluation, including stool examination for occult blood.

VI. LATE PHASE II - PHASE III STUDIES

These studies are intended to continue effectiveness determinations in long-term use, refine dose-response assessment, and evaluate long-term toxicity. Most trials should continue to be blinded and randomized, but they will commonly involve control with a comparison agent, as prolonged placebo treatment is generally unacceptable and unnecessary. Patients will usually be outpatients, selected in accordance with criteria listed in Appendices 1 and 2.

In this phase, in addition to further dose-response information, the effectiveness and tolerance of various dose intervals may be explored. In these long-term studies, concomitant therapy of various kinds will be used by most patients, providing an opportunity to explore drug-drug interactions. Ordinarily, other NSAID therapy should be avoided. If possible, concomitant analgesic or corticosteroid therapy should be kept constant. Alternatively, changes in concomitant treatment can be used as a measure of effectiveness (e.g., steroid or analgesia-sparing effect). In any event, whenever

concomitant therapy is used, care should be taken to record when it is used during the study, and at what dose, and all measurements (baseline and treatment) should be defined with respect to all therapy being taken.

As in other studies, patients should be evaluated at defined intervals. In general, observations should be more frequent early in the study, and can be less frequent later. Baseline observations should be adequate to define initial status and variability.

It may be appropriate late in Phase III to enlarge the exposed patient population, by initiating open studies, or by continuing patients from the controlled studies in open extensions.

DEGENERATIVE JOINT DISEASE

(OSTEOARTHRITIS)

I. INTRODUCTION

A. "General Considerations for the Clinical Evaluation of Drugs", "Guideline for the Format and Content of the Clinical and Statistical Sections of an Application" and the Introduction to the Rheumatoid Arthritis Section of these guidelines should be used for guiding the clinical program.

B. General Comments

Degenerative joint disease is a disorder of moveable joints, characterized by deterioration and abrasion of articular cartilage and also formation of new bone at joint surfaces. Signs and symptoms of inflammation may be present.

Involved joints may be grouped for drug studies as follows:

1. Large joints --

   (a) Hips
   (b) Knees
   (c) Hips and knees
   (d) Ankles
   (e) Shoulders

2. Small joints -- Interphalangeal osteoarthritis should have more than one actively involved joint.

3. Spine

   (a) Lumbar
   (b) Cervical
   (c) Lumbar and cervical

For trials with the primary purpose of investigating the effectiveness of the drug (usually Phase I and II studies), it is advisable to study patients with one (large) target joint, such as the knee or the hip. In Phase III studies, however, where the main objective is the collection of adverse reaction data, a combination of joints from the groupings listed above may be studied.

II. PATIENT SELECTION

In Phases I and II, patients with primary or posttraumatic osteoarthritis should be selected for studies with nonsteroidal anti-inflammatory drugs. The following inclusion criteria should be met:

1. Roentgenological evidence of degenerative joint disease.

2. One or more of the following: swelling, heat, redness, tenderness on pressure, and/or pain on motion or at rest.

   (If any of these are to be used as primary efficacy variables, they should be inclusion criteria for all patients, and only patients with at least moderate baseline severity, on whatever scale is to be used, should be entered into the study.)

   Patients with concomitant disease, that may affect joints, should be excluded. Examples are: psoriasis, syphilitic neuropathy, ochronosis, metabolic bone disease, and acute trauma, with or without degenerative joint disease. Patients with chondrocalcinosis on x-ray, chronic pain syndrome, or drug abuse patterns should be excluded from Phase I and II studies. Phase III studies may include some of these patients.

III. PHASE I STUDIES

See "General Considerations for the Clinical Evaluation of Drugs."

A. Investigators

   Observers experienced in the evaluation of new anti-arthritic compounds (e.g., rheumatologists, orthopedists, or other qualified investigators).

B. Setting and Observations

   Patients may be hospitalized or followed as outpatients, depending upon the safety status and previous experience with the drug. Some pilot studies may be open-label with variable doses. Analgesic drugs should not be added. Duration of treatment should be generally 3 to 6 weeks.

   Safety should be monitored clinically, as well as with a battery of standard laboratory tests and any other laboratory tests suggested by preclinical studies. In early studies, monitoring should be more frequent than in later studies. In rising dose studies, monitoring should also be more frequent.

In addition to assessing primary efficacy measures, efficacy evaluation of other manifestations of disease, even though they may not be shared by all patients, should be carried out with respect to change in swelling, redness, tenderness on pressure, pain at rest or on motion, change in range of motion, and walking or stair-climbing time as supportive evidence of efficacy. In addition, the investigator's and patient's opinion of the patient's condition on the day of assessment should be recorded. Both opinions should be graded descriptively with standardized scales. It is recommended that efficacy be assessed at least weekly.

IV. PHASE II AND III STUDIES

A. Early Phase II

1. Investigators (See Phase I)

2. Setting and Observations

Phase II studies are intended to provide unequivocal evidence of effectiveness from well-controlled studies, and to begin to explore the dose-response curve(s) of the drug.

The studies may be done in outpatients. Analgesic drugs should not be added. These studies may be of a parallel or crossover design, but crossover designs may pose problems that should be anticipated. If a crossover design is used, the drug-free interval at the crossover point should be determined by the half-life of the drug or, alternatively, one can establish a new baseline for the second period by allowing patients to flare, or one can even cross patients over without a washout and not use the data from the first part of the next efficacy period. Theoretically, allowing the patients to flare, and comparing their treatment responses to the appropriate baseline and their baselines to one another, is the best approach. In practice, however, the baselines frequently vary, and it is the rule rather than the exception to see a period effect, so that the theoretical advantage of reestablishing the baseline, or even of using the crossover design, is seldom realized. In any case, the study must be analyzed for the presence of period effects.

Because crossover studies so frequently show a period effect, it is recommended that sufficient patients be enrolled to permit analysis of the 1st period as two parallel groups, if the assumptions for a crossover analysis are not met.

Safety should be monitored clinically as well as with standard laboratory tests and any other tests suggested by preclinical and Phase I studies.

In addition to assessing primary efficacy measures, efficacy evaluation of other manifestations of disease, even though they may not be shared by all patients, should be carried out with respect to change in swelling, redness, tenderness on pressure, pain at rest or on motion, change in range of motion, and walking or stair-climbing time as supportive evidence of efficacy. In addition, the investigator's and patient's opinion of the patient's condition on the day of assessment should be recorded. Both opinions should be graded descriptively with standardized scales. It is recommended that efficacy be assessed at least every 2 weeks.

B. Late Phase II and Phase III

Expanded phases of clinical investigation of this disease continue to be directed toward evaluation of safety and efficacy. Some double-blind controlled studies with active drugs should be included. These studies should be conducted for a period of at least one year, if the drug is to be utilized in the long-term management of degenerative joint disease. In addition, open-label studies, primarily intended to collect data about adverse effects, may also be performed. Parallel design is usually more appropriate for long-term studies.

Criteria for evaluation of efficacy should be as previously described. It is recommended that some studies focus specially on patients with hip, knee, hand, and/or cervical spine, to explore whether or not there is any particular advantage, or disadvantage, in treating patients with degenerative joint disease of different joints.

ACUTE BURSITIS AND TENDINITIS

I. INTRODUCTION

A. Relevant Guidelines

"General Considerations for the Clinical Evaluation of Drugs" as well as the earlier sections of these guidelines should be regarded as the background for this section. In addition, a drug to be used in acute bursitis and tendinitis ordinarily will have been studied in patients with rheumatoid arthritis and/or patients with degenerative joint disease earlier in the drug's development, because those diseases provide a more stable background upon which to evaluate drug effects.

B. General Comments

Acute bursitis or tendinitis may involve any bursa or tendon in the body, but should involve only one area at the time of the study. A drug investigation program may include any or all of the following:

1. Acute shoulder --

   (a) Bicipital tendinitis

   (b) Subdeltoid bursitis

2. Olecranon bursitis

3. Lateral epicondylitis

Difficulty in differentiating between bicipital tendinitis and subdeltoid bursitis is recognized. Therefore, the term "acute shoulder" is utilized to include involvement of the tendon and/or bursa.

For drug studies, each major grouping should be evaluated separately. Subgrouping may be utilized in controlled studies, but this should be predefined.

II. PATIENT SELECTION

It is assumed that the drug is being studied in rheumatoid arthritis and/or osteoarthritis, so that clinical pharmacology and overall safety can be evaluated primarily in those trials. Moreover, Phase I and II studies in those conditions should provide a reasonable estimate of the dose to be used in these studies.

Each patient should fulfill all of the following criteria for acceptance into a study:

1. An acute episode of less than 4 but, preferably, less than 3 days duration. (The longer the duration of acute symptoms prior to therapy, the greater the risk of failing to show a significant difference between drug and placebo, or the greater the number of patients needed in a positive control study.) There should be a history of painless unrestricted motion of the affected joint immediately prior to the acute attack.

2. Localized tenderness over the involved area.

3. Limitation of motion.

4. Pain on motion.

5. Pain sufficiently severe to interfere with sleep.

The following conditions should be excluded with appropriate tests including, as a minimum, a roentgenogram of the affected area and an erythrocyte sedimentation rate:

1. Fracture.

2. Rheumatoid arthritis, ankylosing spondylitis, or other connective tissue disease.

3. Crystal-induced arthropathy, i.e., gout, pseudogout. (Exclusion may be based on history or current findings.)

4. Degenerative joint disease.

5. Shoulder-hand syndrome.

6. Cervical radiculopathy.

7. Specific infectious processes.

8. Malignancy.

III. PHASE I to III STUDIES

See "General Considerations for the Clinical Evaluation of Drugs."

A. Investigators

Rheumatologists and other qualified individuals experienced in the evaluation of new anti-inflammatory compounds. It is recommended

that the principal investigator be a rheumatologist, or an orthopedist, experienced with rheumatic disease patients as well as acute trauma.

B. Setting and Observations

Patients will ordinarily not be hospitalized for these conditions, and initial dose finding and safety assessments will be available. Initial pilot studies may be performed open-labeled with variable doses, but the principal objective will be to carry out controlled studies to demonstrate effectiveness in these conditions.

Double-blind efficacy studies of parallel design with the test drug and a placebo may be conducted at low, medium, and high doses to obtain a comparison of efficacy at different dose levels. For relatively long half-life drugs, when total body accumulation of a drug is greater than two times the usual single dose, loading doses should be investigated. In other studies, efficacy may be evaluated in double-blind, randomized comparison with other nonsteroidal anti-inflammatory agents of established efficacy.

Non-anti-inflammatory analgesics (e.g., codeine) may be administered as rescue medicine during the first 48 hours under predefined conditions, with arrangements for their use and the careful recording of their consumption specified in the protocol. Their use may be included as a measure of efficacy. Rest, application of heat or ice, and physical therapy may be utilized, if they are carefully standardized. It is particularly important to record these forms of therapy in double-blind studies.

Efficacy should be evaluated with respect to change in swelling, redness, tenderness on pressure, pain at rest, pain on motion or range of motion, both active and passive. The use of rescue medication, and the handling of other efficacy outcome variables after its use, should be defined prior to the study, and the potential effects of its use on the study's results should be specifically addressed in the study's analysis. The investigator's and patient's opinion of the patient's response to treatment should be obtained. Both opinions should be descriptively graded, and this assessment should be done with consistency (i.e., very good, good, fair, poor, very poor).

Duration of treatment should not exceed 14 days. Evaluation should be performed at least on days 0, 1, 2, 4, 7, and 14 of drug treatment, unless the new drug is stopped. Multiple evaluations during the first 48 to 72 hours are encouraged. Post-treatment evaluations of outcome, and for any delayed adverse events, should be performed at 2 to 4 weeks.

ANKYLOSING SPONDYLITIS

I. INTRODUCTION

"General Considerations for the Clinical Evaluation of Drugs" as well as the Introduction to the Adult Section of these guidelines should be used for guiding the clinical program. In addition, a drug to be used in ankylosing spondylitis ordinarily will have been studied in patients with rheumatoid arthritis earlier in the drug's development because of the greater number of patients with rheumatoid arthritis.

II. PATIENT SELECTION

All patients should have a definite diagnosis of active ankylosing spondylitis, as defined in Appendix 4, and active disease at the time of entering into any drug studies. This activity may be demonstrated by withdrawing drugs from patients receiving therapy. However, patients with active disease, only partially controlled by other drugs, may remain on other drugs with the addition of the test drug for Phase II and III studies. The dosage of the background drugs should remain constant and usage of these drugs should be predefined.

Patients should be excluded who: (1) are less than 16 years of age; (2) have active involvement of more than two peripheral joints, not including the shoulders or hips; or (3) have spondylitis associated with psoriatic arthritis, Reiter's Syndrome, or inflammatory bowel disease.

III. PHASE I to III STUDIES

See "General Considerations for the Clinical Evaluation of Drugs."

A. Investigators

Rheumatologists and other qualified individuals experienced in the evaluation of new anti-inflammatory compounds. It is recommended that the principal investigator be a rheumatologist.

B. Setting and Observations

It is presumed that the drug is being studied in rheumatoid arthritis so that clinical pharmacology and overall safety can be evaluated

primarily from those trials. Moreover, Phase I and II studies in rheumatoid arthritis should provide a reasonable estimate of the dose to be used in these studies.

Patients may be hospitalized, or outpatients, depending upon the safety status of previous experience with the drug under evaluation. They should be followed closely for evidence of adverse reactions by clinical observations and routine safety laboratory tests, and any other laboratory tests indicated by preclinical animal studies or previous clinical experience.

Initially, open-label studies may be conducted, of 4 to 6 weeks duration, with variable doses. Measurements for assessing the response of ankylosing spondylitis to drugs (Appendix 5) should be recorded at the time of performance. In early studies, patients should be assessed at least weekly for efficacy and safety.

Some short-term studies may be crossover in design, randomized, and double-blind, conducted with (a) placebo and the test drug at low, medium, and high doses for comparison of efficacy; (b) placebo and test drug at a fixed dose; (c) test drug and an adequate dose of a nonsteroidal anti-inflammatory drug of established efficacy. There should be a washout period between treatments, during which a baseline is established for the treatment that follows and for comparison with the other baseline period(s). Because the number of patients with ankylosing spondylitis is limited and because the efficacy outcome variables seem to change more slowly than those in rheumatoid arthritis and osteoarthritis, crossover designs have been used more in studying treatment of this indication. One needs to be aware of all the problems as discussed in the Osteoarthritis Section and, as a practical matter, one is still well advised to design the study to have at least the possibility of sufficient power to analyze the first period as a parallel group study.

Other studies should be of parallel design, and conducted for 6 to 12 weeks, using the same assessment of response as above (Appendix 5).

If necessary for the patient's comfort, a non-anti-inflammatory analgesic may be added. However, the reasons for so doing should be specified, and only one specific analgesic drug should be used for the patients in a given study. No analgesic drug combinations should be employed. The planned analysis should include provisions for dealing with additional analgesics. The final analysis should explore the potential effect on the conclusions of the planned and actual analyses, if they were different. The patient's physical therapy program should not be modified during the study period. If a change in a patient's physical therapy program occurs, its potential impact on the outcome should be discussed.

In later trials it may be appropriate to involve patients who have spondylitis who do not meet the inclusion criteria above, particularly in the case of patients with early disease, with definite sacroiliac joint involvement, or patients with spondylitis associated with psoriatic arthritis, Reiter's Syndrome, inflammatory small bowel disease, or patients with definite spondylitis and more than two peripheral joints involved. The purpose of these studies is not to prove effectiveness but rather to screen primarily for any special problems that might be associated with the new drug in these smaller subsets of patients with other diseases who are perhaps on other drugs as well.

ACUTE GOUTY ARTHRITIS

I. INTRODUCTION

A. Relevant Guidelines

"General Considerations for the Clinical Evaluation of Drugs" as well as the Introduction to the Rheumatoid Arthritis Section of these guidelines should be used for guiding the clinical program. This guideline for acute gouty arthritis assumes, as do the previous two, that the drug has already been studied in patients with rheumatoid arthritis and/or osteoarthritis. There have been a few nonsteroidal anti-inflammatory drugs with specific effects on uric acid kinetics, which has initially led to their study in patients with gout. In such cases, Phase I and II studies need to be performed following the general guidelines for early studies, as outlined in the Rheumatoid Arthritis and Osteoarthritis Sections.

B. General Comments

A diagnosis of gout requires fulfilling of the following criteria:

1. Hyperuricemia.

2. Either (a) or (b), below. [If (b) is used, at least two of the four conditions should be fulfilled.]

   (a) Evidence of monosodium urate crystals in the synovial fluid.

   (b) History --

   (1) A clear history (and/or observation) of at least two attacks of painful, limb joint swelling. These attacks, at least in the early stages, must exhibit an abrupt onset of severe pain and complete remission within a week or two.

   (2) A clear history (and/or observation) of podagra: that is, an attack, as described above, involving the great toe.

   (3) The presence of tophi.

   (4) A clear history (and/or observation) of a good response to colchicine, defined as a major reduction in objective signs of inflammation within 48 hours of the onset of therapy.

II. PATIENT SELECTION

Patients with acute gout who are selected for drug studies should have the following:

1. Acute involvement of one or two joints, usually the lower extremity, of less than 48 hours duration.

2. History of an abrupt onset of significant pain in the involved joint, accompanied by tenderness, redness, and heat.

3. Exclusions: any other rheumatic disorder that might interfere with the assessment or interpretation of response.

III. PHASE I to III STUDIES

A. Investigators

Rheumatologists and other qualified investigators experienced in the evaluation of anti-arthritic compounds. It is recommended that the principal investigator be a rheumatologist or orthopedist experienced in treating patients with rheumatic disease as well as acute trauma.

B. Setting and Observations

Studies may be performed in outpatients or hospitalized patients. The effect of the test drug on change in pain, swelling, tenderness, and range of motion should be descriptively graded with consistency (for example, mild, moderate, and severe). The investigator's subjective opinion and the patient's opinion should likewise be recorded on a scale that utilizes consistent terminology throughout the program. Assessments should be recorded prior to therapy, and at least once, but preferably more times daily for at least 3 days during treatment. At least one assessment should be performed several days after completing the treatment. A good or an excellent response in outcome in a patient may be defined as improvement by two-thirds or better in efficacy measurements within a 72-hour period. Some studies, using doses demonstrated to be effective in other studies or conditions, may be performed in outpatients as short-term, 5 to 10 day, controlled, open-label studies, using historical controls, because the course of untreated gout is sufficiently predictable by outcome variables that do not require blinding. The studies require formal analysis and comparison with series of untreated cases from the literature. Additional studies may be of parallel design, with double-blind efficacy comparison of the test drug with appropriate doses of either colchicine, or an approved NSAID, such as indomethacin, of established efficacy in

treating acute gouty arthritis. Studies at low, medium, and high drug doses should be performed for comparison of efficacy at different dose levels, because patients with gout frequently require higher doses of some drugs. The use of non-anti-inflammatory analgesic drugs and physiotherapy should be recorded in open-label studies, and be carefully predefined and monitored in controlled studies. Patients should be monitored at close intervals for adverse effects, with clinical and laboratory examinations.

The effect of the new drug on uric acid metabolism, and the concomitant use of other agents (such as allopurinol or probenecid), which alter uric acid metabolism, should be evaluated.

INTERVAL OR INTERCRITICAL PHASE GOUT AND CHRONIC GOUT

GENERAL COMMENTS

Studies on interval or intercritical phase gout may be conducted with the test drug. Controlled studies of parallel design should be included. In some of these, a nonsteroidal anti-inflammatory drug of established efficacy may be administered in a comparable dose, based on other studies. The treatment and control groups should be comparable with respect to prestudy severity and incidence of attacks. Some studies should be performed with the new drug plus drugs known to affect uric acid metabolism. Studies should be conducted with the new drug in interval gout of at least one year's duration, in more than 100 patients, in order to evaluate adverse effects and to estimate the reduction in severity and frequency of attacks, if the drug is to be recommended for chronic use.

PEDIATRIC SECTION

JUVENILE RHEUMATOID ARTHRITIS (JRA)

I. INTRODUCTION

A. Relevant Guidelines

"General Considerations for the Clinical Evaluation of Drugs" and "General Considerations for the Clinical Evaluation of Drugs in Infants and Children" should be reviewed before reading these guidelines.

B. General Comments

Before a new nonsteroidal anti-inflammatory agent is tested in children, information on drug distribution and information about the effectiveness and adverse effects in the dosage range to be studied in children should be available from studies in adults with appropriate diseases (see 7 below).

Several unique characteristics of children and pediatric practice need emphasis:

1. The child is growing in weight and height.

2. Tissues and organ systems of the child are growing and developing.

3. Intellectual and social functions of the child are developing.

4. Sexual development and future reproductive capacity of children may be altered by exogenous agents.

5. A long time period exists between early childhood and the time of full growth and development.

6. Because of differences in weight and length during the growth period, some drugs may best be administered on a surface area basis rather than on a weight basis.

7. Diseases of children may differ considerably in manifestations and prognosis from counterpart diseases of adults (for example, juvenile rheumatoid arthritis and adult onset rheumatoid arthritis). Some disease states common in adulthood exist only rarely in childhood (for example, degenerative joint disease, gout, bursitis syndromes).

8. Use of a drug in childhood allows for a much longer period of time in which to manifest slow appearing conditions, such as malignancies. These conditions may not appear until many years later.

   Therefore, chronic administration of new drugs in children must be done with cognizance of possible effects on growth and development of the whole child and his or her parts. Ideally, drugs should be monitored for possible latent or long-term side effects. For this reason, it is strongly recommended that patients in study protocols of DMARDs be informed of the need for long-term follow-up, and that a list of these patients' names, status, and locations be maintained by the sponsors, pending the development of guidelines for registries. Pharmaceutical companies are encouraged to take responsibility for long-term pre- and post-marketing surveillance. Emphasis should be placed on accelerating accumulation of knowledge about each newly developed and newly marketed drug.

II. INVESTIGATORS

Investigators for studies in children should be physicians experienced in the care of children, in clinical trials, and in pediatric rheumatology.

III. SETTINGS

Most studies will be under outpatient circumstances. Adequate laboratory and hospital facilities for children must be available, and the amounts of blood required for testing should be minimized.

IV. TYPES OF PATIENTS

The diagnosis of JRA must be established, and patients should be categorized into the appropriate clinical patterns (Appendix 6). Patients should be no younger than 18 months, and no older than 17 years, to be enrolled in a study. Since both the natural history of disease and drug efficacy or toxicity may differ for the various types of JRA, care should be taken to balance the allocation of medications within each clinical pattern, and an analysis should be performed stratifying the patients by disease type.

On admission to study, patients should have active disease (Appendix 6). Ideally, patients should have disease of recent onset, either untreated or treated only with a nonsteroidal anti-inflammatory agent. No therapy with gold, glucocorticoids, or chloroquine should have been received within 3 months of study. Studies of nonsteroidal anti-inflammatory agents in patients continuing to receive gold, chloroquine, or glucocorticoid therapy would require different designs.

V. TYPES OF STUDIES

A. Description

The following types of studies should be carried out, generally in the order they are discussed below.

1. Open-label, short-term: To verify safe and optimal doses for children, to insure that short-term toxicity in children is not excessive. Duration: 4 to 8 weeks. Blood levels of the drug should be spot-checked to determine whether or not drug distribution parameters in children are similar to those previously determined in adults.

2. If, in the short-term open trials, effectiveness in children seems comparable to that in adults, then longer term open-label studies to confirm the comparability of adverse effects should be conducted in 75 to 100 children, for periods of 1 or more years. Controlled double-blind comparisons with aspirin or another NSAID approved for use in children are necessary only if the open-label experience suggests some difference between adults and children in the effectiveness or adverse effects of the drug. Sample sizes and duration of such studies will depend on the hypothesis being tested (see the general comments about clinical studies at the beginning of the guidelines).

B. Study and Comparison of Drug Dosages

Prior adult studies will suggest appropriate initial dosages for children. Children's doses can initially be derived by adjusting adult doses on the basis of body weight or body surface area, but the appropriateness and accuracy of these predictions should be checked by determination of blood levels.

Aspirin, or any NSAID approved for JRA, is a suitable comparison drug for double-blind trials. Double-blind comparative trials in children are only required if the results from the open-label experience suggests to the pediatric rheumatologists doing the open trials either some advantage, or special problem, with the new drug's use in children. Doses of the control drug for such studies may be determined in advance for each patient or, if applicable to the two drugs under study, a system of blinded dosage titration should be part of the study design, so that the maximum safe dose of either drug for the individual patient is not exceeded in a blinded fashion, yet adequate amounts are being used. Spot checking of blood levels or amounts excreted in the urine, to be monitored by a blinded observer, is recommended.

C. Observations

1. Frequency

Observations should be made weekly for the first 4 to 6 weeks of study, and at least monthly thereafter during the first year of study.

2. Effectiveness

In general, effectiveness is evaluated as described in Appendix 3. Note should be taken of effects of drugs on the different clinical manifestations of JRA, such as systemic disease, arthritis, and iridocyclitis. Details of all dropouts, and the causes thereof, should be recorded. Although abnormalities due to disease may not be demonstrable by radiography early in the disease course, involved joints should be radiographed before and after any study period exceeding 6 months. The American Rheumatism Association's Council on Pediatric Rheumatology has developed a functional classification applicable to children with JRA. It is recommended that it be used, although there is not sufficient experience with it to date to justify its use as a primary efficacy variable (see the general comments on clinical trial design and analysis at the beginning of the guidelines).

3. Safety

Adverse effects of drug administration should be evaluated not only in specific organ systems, as in adults, but also with respect to growth, development, and behavior. Ophthalmologic and hearing assessments should be made prior to study and at appropriate intervals during the trial.

4. Need for Long-Term Studies

Possible effects on maturation, reproductive capacity, and other areas may not become apparent until many years after drug administration. Unfortunately, no clear methods for detecting such events in any systematic fashion are now available. Long-term follow-up, coupled with automated data handling of clinical information, may permit such studies in the future.

APPENDIX 1

CRITERIA FOR THE DIAGNOSIS OF RHEUMATOID ARTHRITIS

CRITERIA FOR THE EXCLUSION AND SELECTION OF SUITABLE PATIENTS FOR NEW DRUG INVESTIGATION

The criteria listed below have been taken from Reference 1 by Ropes et al. The final paragraph lists additional recommended exclusions.

DIAGNOSIS

The diagnosis requires five of the following criteria. In criteria 1 through 5 the joint signs or symptoms must be continuous for at least 6 weeks. (Any one of the features listed under EXCLUSIONS will disqualify a patient from being diagnosed as having rheumatoid arthritis.)

1. Morning stiffness.

2. Pain on motion or tenderness in at least one joint (observed by physician).

3. Swelling (soft tissue thickening or fluid - not bony overgrowth alone) in at least one joint (observed by physician).

4. Swelling (observed by physician) of at least one other joint. (Any interval free of joint symptoms between the two joint involvements may not be more than 3 months in duration.)

5. Symmetrical joint swelling (observed by physician) with simultaneous involvement of the same joint on both sides of the body. (Bilateral involvement of midphalangeal, metacarpophalangeal, or metatarso-phalangeal joints is acceptable without absolute symmetry.) Terminal phalangeal joint involvement will not satisfy this criterion.

6. Subcutaneous nodules (observed by physician) over bone prominences on extensor surfaces or in juxta-articular regions.

7. X-ray changes typical of rheumatoid arthritis (which must include at least bony decalcification localized to, or greatest around, the involved joints, and not just degenerative changes). Degenerative changes do not exclude patients from any group classified as rheumatoid arthritis.

8. Demonstration of the "rheumatoid factor" by any method that, in two laboratories, has been positive in not over 5 percent of normal controls.

9. Poor mucin precipitate from synovial fluid (with shreds and cloudy solution).

10. Characteristic histological changes in synovial membrane with three or more of the following: marked villous hypertrophy; proliferation of synovial lining cells; marked infiltration of chronic inflammatory cells (lymphocytes or plasma cells predominating) with tendency to form "lymphoid nodules," deposition of compact fibrin either on surface or interstitially; foci of cell necrosis.

11. Characteristic histological changes in nodules, showing granulomatous foci with central zones of cell necrosis, surrounded by proliferated fixed cells, peripheral fibrosis, and chronic inflammatory cell infiltration, predominantly perivascular.

EXCLUSIONS*

1. The typical rash of disseminated lupus erythematosus (with butterfly distribution, follicle plugging, and areas of atrophy).

2. High concentration of lupus erythematosus cells (four or more in two smears, prepared from heparinized blood, incubated not over 2 hours).

3. Histologic evidence of periarteritis nodosa, with segmental necrosis of arteries, associated with nodular leucocytic infiltration, extending perivascularly and tending to include many eosinophils.

4. Weakness of neck, trunk, and pharyngeal muscles, and persistent muscle swelling of dermatomyositis.

5. Definite scleroderma (not limited to the fingers).

6. A clinical picture characteristic of rheumatic fever, with migratory joint involvement and evidence of endocarditis, especially if accompanied by subcutaneous nodules, or erythema marginatum, or chorea. (An elevated antistreptolysin titer will not rule out the diagnosis of rheumatoid arthritis.)

---

* It is obvious that complete documentation for these exclusions is not necessary, but the listed entities should be considered.

7. A clinical picture characteristic of gouty arthritis, with acute attacks of swelling, redness, and pain in one or more joints, especially if relieved by colchicine.

8. Tophi.

9. A clinical picture characteristic of acute infectious arthritis, of bacterial or viral origin, with an acute focus of infection, or in close association with a disease of known infectious origin.

10. Tubercule bacilli in joints, or histologic evidence of joint tuberculosis.

11. A clinical picture characteristic of Reiter's syndrome, with urethritis and conjunctivitis associated with acute joint involvement, usually migratory initially.

12. A clinical picture characteristic of the shoulder-hand syndrome: unilateral involvement of shoulder and hand, with diffuse swelling of the hand followed by atrophy and contractures (25% of shoulder-hand syndromes are bilateral).

13. A clinical picture characteristic of hypertrophic pulmonary osteoarthropathy, with clubbing of fingers and/or hypertrophic periostitis, along with shafts of the long bones, especially if an intrapulmonary lesion is present.

14. A clinical picture characteristic of neuroarthropathy, with condensation and destruction of bones of involved joints and with associated neurologic findings.

15. Homogentisic acid in the urine, detectable grossly with alkalinization (ochronosis).

16. Histologic evidence of sarcoid or positive Kveim test.

17. Multiple myeloma, as evidenced by marked increase in plasma cells in the bone marrow, or Bence-Jones protein in the urine.

18. Characteristic skin lesions of erythema nodosum.

19. Leukemia or lymphoma, with characteristic cells in peripheral blood, bone marrow, or tissue.

20. Agammaglobulinemia.

In addition, patients with ankylosing spondylitis, rheumatoid arthritis with onset prior to age 16, and patients with psoriasis or inflammatory bowel disease should be excluded from drug trials of "rheumatoid arthritis."

APPENDIX 2

RHEUMATOID ARTHRITIS: DEFINITION OF ACTIVE DISEASE

Active disease is defined by showing at least three of the following (Reference 2)

1. Number of joints tender or painful on motion: 6 or more.
2. Number of joints swollen: 3 or more.
3 Duration of morning stiffness (hours): 3/4 or more.
4. Westergren sedimentation rate (mm/hr): 28 or more.

APPENDIX 3

RHEUMATOID ARTHRITIS: EFFICACY ASSESSMENT (Reference 2)

A. All Studies*

1. Number of painful or tender joints

   Sixty-eight (68) joints are assessed. To be included are: temporomandibular (2); sternoclavicular (2); and acromioclavicular (2); radiocarpal, carpal, and carpometacarpal are collectively designated wrist (2). Ankle (2) is the mortise. Tarsus (2) includes as a single unit subtalar, transverse tarsal, and tarsometatarsal joints. The eight IP joints (proximal and distal) of the four lateral toes are counted as four units.

   Press on joint and move through a full range of motion. Pain upon either maneuver represents a positive result. Record which joints are positive.

2. Number of swollen joints

   Sixty-six (66) joints are assessed. The hips are omitted from the above list. Synovial fluid and/or soft tissue swelling, but not bony overgrowth, represents a positive result. Record which joints are positive.

3. Duration of morning stiffness

   Ask the patient to think of recent mornings. When did the patient awake? When did the sensation of stiffness begin to wear off?

4. Grip strength

   It is best to use a sphygmomanometer cuff, which has been sewn into a cloth bag that fits snugly into the palm after folding the cuff twice and thus reducing its dimensions to 1/3 of the original. Inflate to 20 mm Hg. With arm unsupported, the

---

* Items 1, 2, 7, and 8 are considered primary efficacy variables and items 3 to 6 are regarded as secondary. It is recommended that all eight be recorded and analyzed in all studies.

patient should grasp and squeeze three times. Note the sustained values. The procedure is repeated for the other hand. The six values are averaged and the mean is recorded.

5. Time required to walk 50 feet

   Specify what walking aids are required. Use a stop watch to record the time required to walk (not run) a marked 50-foot distance from a standing start. Record to the nearest tenth of a second.

6. Erythrocyte sedimentation rate (recommend only the Westergren method)

   Two ml of whole blood (drawn with tourniquet removed) added to 0.5 ml of 3.8% sodium citrate. The mixture is drawn up to 200 mm in a 300 x 2.5 mm tube and allowed to stand precisely vertical for 1 hour.

   The clear zone over the red cells is then measured in millimeters.

7. Principal observer's opinion of patient's condition on day of assessment. This assessment should be made and recorded before obtaining the patient's opinion. This opinion should be descriptively graded and with consistency (i.e., very good, good, fair, poor, very poor).

8. Patient's opinion of condition on day of assessment. This should also be descriptively graded and with consistency.

B. Chronic Studies

For studies of 6 months or longer, four further modes may be considered in addition to the above:

1. Record ARA (Steinbrocker) functional capacity (class), as described in Reference 3, before, during, and at end of drug trial. Although this is highly recommended for chronic studies, it may also be of value in studies of less than 6 months duration.

2. ARA anatomical stage, as described in Reference 3, recorded before and at end of drug trial.

3. Of supportive importance: hand x-ray made at beginning and end of the trial, without knowledge of drugs or clinical course.

4. Of secondary importance: serum samples from beginning and end of trial to be analyzed for rheumatoid factor titer values (parallel sample testing) without knowledge of which serum is first or last, or information on drugs or clinical course.

APPENDIX 4

I. Criteria for Definite Diagnosis of Ankylosing Spondylitis

A. Clinical Criteria

1. Lumbar or dorsal lumbar junction pain, and stiffness of over 3 months duration.

2. Major limitation of motion of lumbar spine in three directions: Flexion-extension, lateral bending and rotation (see Appendix 5).

3. Pain and stiffness in the thoracic region of over 3 months duration.

4. Limited chest expansion (for build, age, weight: generally less than 5 cm, but may be quite variable).

5. Nocturnal pain with morning predominance, involving the axial skeleton, and/or morning stiffness, and/or bilateral pain in buttocks, or pain in either buttock.

B. X-Ray Criteria: Grading of the Sacroiliac Joints as Follows:

0 = Normal

1 = Suspicious

2 = Definitely abnormal

3 = Advanced abnormal; includes some ankylosis

C. A Diagnosis of Definite Ankylosing Spondylitis Required:

1. Grade 2 or 3 bilateral sacroiliiitis by x-ray.

2. Definite history of at least two clinical criteria.

II. Definition of Active Spondylitis

Requires spinal and/or sacroiliac pain, and one or more of the following:

1. Morning stiffness of greater than 1/2 hour duration.

2. Active involvement of one or more peripheral joints (including shoulders or hips).

3. Increased sedimentation rate.

APPENDIX 5

MEASUREMENTS FOR ASSESSING THE RESPONSE OF ANKYLOSING SPONDYLITIS TO DRUGS

1. Objective Evaluation by Investigator of Spinal Pain

   Tenderness and/or limitation of motion in three areas of the spine: cervical, thoracic, and lumbar plus sacroiliacs. Limitation of motion refers to that caused by pain and spasm over and above any permanent mechanical limitation. These findings should be recorded in a range from 0 to 4 as follows:

   0 = No pain on firm palpation, percussion, and on extreme motion. No spasm.

   1 = Slight pain on firm palpation, percussion, and on extreme motion. No more than slight limitation of motion.

   2 = Moderate pain on moderate palpation, percussion, or motion. No more than slight limitation of motion.

   3 = Moderate to severe pain on light palpation, percussion, or slight motion. Moderate to severe limitation of motion.

   4 = Extreme pain with inability to withstand even light palpation or percussion, and essentially no mobility of spine.

2. Subjective Evaluation by the Patient

   Spinal pain is to be evaluated and recorded for cervical, thoracic, and the lumbar/sacroiliacs, using a scale from 0 to 4, as follows:

   0 = No pain.

   1 = Slight steady pain, and/or only an occasional twinge, on extreme motion or stress.

   2 = Moderate steady pain, or slight steady pain, with frequent moderate to severe episodes.

   3 = Severe steady pain, or moderate to severe steady pain, with frequent very severe episodes.

   4 = Extremely severe pain ("worse possible") which is constant and totally disabling.

3. Morning stiffness

   The patient should be instructed that the duration of morning stiffness is the elapsed time between arising in the morning and the development of noticeable improvement in the degree of stiffness. This is recorded in hours to the nearest half-hour.

4. Number of peripheral joints (including shoulders and hips) involved by swelling, pain, or tenderness.

5. Erythrocyte Sedimentation Rate (Westergren Method)

6. Chest Expansion

   Chest expansion should be measured as the difference in centimeters between full expiration and full inspiration, measured at the nipple line. The better of two tries should be recorded.

7. Fingers-to-Floor Test

   This test is performed with knees straight as possible, with the distance in centimeters measured from extended fingertips to the floor, with maximum effort to touch the floor. The best of two tries should be recorded.

8. Occiput-to-Wall Test

   This test is performed with the heels, if possible, and the back against the wall, or the edge of a door, with the distance measured in centimeters from the occiput to the wall during maximum effort to touch the head to the wall, without raising the chin above its usual carrying level. The best of two tries should be recorded.

9. Schober Test

   This test is performed by marking a point over the spinous process of L-5 (found as the first process below the projected line across the back at the level of the top of the iliac crest, with exact L-5 not too critical to the test). The second point is measured 10 centimeters directly above the first while the patient is hyperextending his lumbar spine as much as possible. The patient then flexes forward maximally and the (now longer) distance between the two points is measured. Normally the initial 10 cm distance increases to 16 cm or more. The change should be written ("10/16") in this case, while 10/10 would indicate no change on flexion of the lumbar spine. For recording this measurement, write in the actual distance in

centimeters obtained on the measurement made in full flexion (not the difference between the extension and flexion measurements).

10. Night Pain

This evaluation of night pain encompasses frequency, severity, sleep loss, etc., and should be recorded as follows:

1 = Not bothered; no pain at all.

2 = Bothered a little; pain is present part of the time, but dull in character.

3 = Bothered a lot; steady or intermittent pain, which usually interferes with sleep.

4 = Bothered terribly; the night pain is constant, causes marked interference with sleep, and the patient is quite miserable.

11. Principal observer's subjective opinion of patient's general condition on day of assessment. This assessment should be made and recorded before obtaining the patient's opinion. This opinion should be descriptively graded and with consistency (i.e., very good, good, fair, very poor).

12. Patient's opinion of general condition on day of assessment. This should also be descriptively graded and with consistency.

APPENDIX 6

JUVENILE RHEUMATOID ARTHRITIS (JRA)

"Guidelines for the Clinical Evaluation of Analgesic Drugs" is an important companion document and should be reviewed prior to reading this Appendix.

The relationship of JRA to adult onset rheumatoid arthritis remains uncertain. JRA differs from adult disease in several aspects, including occurrence of arthritis frequently limited to only a few joints, predominance of systemic symptoms in some patients, association with chronic iridocyclitis, infrequent occurrence of rheumatoid factor and rheumatoid nodules, occurrence of secondary growth disturbances, and better general prognosis.

1. Basic Diagnostic Criteria for JRA (Reference 5)

   a. Onset at or before age 16 years.

   b. Chronic disease with manifestations of either persistent arthritis or typical systemic disease and arthritis of more than 6 consecutive weeks' duration.

   c. Exclusion of other diseases associated with arthritis, including infections, acute rheumatic fever, systemic lupus erythematosus, and ankylosing spondylitis.

2. Classification of Clinical Patterns of JRA

   Children with rheumatoid arthritis fall into one of three broad categories; these disease types are defined by disease manifestations occurring within the first 6 months of disease.

   a. Systemic onset: Prominent systemic manifestations of high intermittent fever to 103 degrees F or greater and rheumatoid rash, with or without hepatosplenomegaly, lymphadenopathy, and polyserositis. Nearly all patients eventually have polyarthritis. Iridocyclitis is very rare.

   b. Polyarticular Disease: Polyarthritis, often symmetrical in distribution, characteristically involves small joints of hands and feet; wrist, knees, ankles, elbows, hips, neck, and jaw are also commonly affected. Mild hepatosplenomegaly and lymphadenopathy may occur. Iridocyclitis is uncommon. Some patients, most often girls, with onset of disease during the teenage years have positive tests for rheumatoid factor and rheumatoid nodules; these patients frequently have severe arthritis.

c. Pauciarticular Disease: Arthritis, often asymmetrical, involves few joints (four or fewer joints by official ARA criteria). Large joints (particularly knees, ankles, and elbows) are more frequently affected than small joints; hip involvement is rare. Chronic iridocyclitis occurs commonly (about 30% of patients); other extra-articular manifestations are rare. Boys, with arthritis onset at an older age (9 years or older), have a significantly higher incidence of HLA B-27 positivity and may represent a different subgroup.

3. Classification of Activity of Disease

   a. Active systemic disease (fever, rheumatoid rash, or other manifestations of systemic-onset JRA).

   b. Active arthritis (objective inflammation in one or more joints).

4. Serologic Studies

   Rheumatoid factors are found in only 10 to 20% of patients, generally those with onset of severe polyarthritis in teenage years.

   Antinuclear antibodies are found in about 30% of patients.

APPENDIX 7

RECOMMENDATIONS FOR DEVELOPING AND SUBMITTING NDAs FOR FIXED-DOSE COMBINATIONS OF NSAIDs WITH CODEINE

In response to requests from sponsors for guidance in developing and submitting NDAs for fixed-dose combinations of marketed nonsteroidal anti-inflammatory drugs (NSAIDs) with codeine for general analgesia indications, FDA has developed some general recommendations regarding the preclinical and clinical portions of such applications.

Preclinical requirements for an NDA for an already marketed NSAID when combined with codeine are (1) the standard preclinical studies for the NSAID and (2) teratology studies of the combination in two species.

Clinical requirements for these NDAs include (1) biopharmaceutical studies to define the rate and extent of absorption for each component of the combination with sufficient power to show a 25% change in either rate or extent of absorption, (2) single dose safety/efficacy studies, and (3) multiple-dose safety/efficacy studies.

In FDA's opinion, the requirements for approval of a combination vary depending on whether the indication is mild to moderate pain, i.e., where it might be possible to enhance the speed of the onset of relief of a drug with a long half life by combination with codeine; moderately severe pain, i.e., where the combination may achieve satisfactory efficacy in painful conditions not adequately treated by either agent by itself; or severe pain, where a combination might be able to achieve pain relief comparable to the injectable narcotics that are normally required. Although in preliminary studies a sponsor may experiment with several pain models using their combination to determine for which level of pain relief it is most suited, in an NDA, results from studies of different pain levels should be separated because the evidence needed to support effectiveness in each is different.

FDA's experience with studies of analgesics has shown that there are more "methodological failures" than is usual in studies of drugs for indications with more objective efficacy endpoints. It is extremely helpful in interpreting a "negative study" to know whether a positive control also was negative in that trial, suggesting a methodological failure rather than lack of effectiveness. These guidelines therefore recommend multiple control groups so that "failed studies" can be more clearly labeled as such and appropriately down weighted. In some cases, the internal consistency provided by these controls will allow fewer studies to be considered adequate support for the claims proposed.

In studies of mild to moderate pain, the fundamental requirement under the combination policy is that the combination must be more effective than each of its components. FDA also recommends inclusion of a placebo group to provide an indication of assay sensitivity and the appropriateness of the population. The NSAID would be expected to be superior to placebo and failure for it to be so would indicate poor sensitivity.

In moderately severe to severe pain where placebo treatment groups are often not feasible, assay sensitivity is ordinarily assessed by using more than one dose of an effective agent. In the case of the combination, FDA recommends inclusion of aspirin and aspirin plus codeine groups, again to provide an indication of assay sensitivity.

Single-Dose Clinical Studies

A. Development of a fixed combination of an NSAID with codeine for a "mild to moderate" pain indication requires evidence of efficacy in each of two pain models (of the sponsor's choosing) comparing the treatment groups shown below. Since the efficacy of the NSAID and codeine will have been previously established, FDA will require only one definitive study in each model to provide evidence of efficacy. This assumes that in studies not showing efficacy, the explanation for the failures will be reasonably apparent from the control groups.

   1. Combination product.

   2. NSAID at the combination dose.

   3. Codeine at the combination dose.

   4. Placebo.

   The combination must be shown to be superior to each component and the NSAID must be superior to placebo in order for the study to be persuasive. A study in which the NSAID's effectiveness was not seen, yet where the combination was superior to its components, might provide some support, but such a peculiar result would weaken it.

B. For a "moderately severe" pain indication two studies will be required in a single pain model, or one study in each of two models, that show the superiority of the combination to each of its components. In addition, the study should be able to distinguish between acetylsalicylic acid (aspirin) 650 mg or acetaminophen 650 mg alone and the combination of one of those drugs with codeine 60 mg; and the test combination should not be less effective than the control combination.

Again it is assumed that the reason for negative results can be identified from the results in the control groups in any studies failing to show efficacy. The studies should have the following treatment groups:

1. Combination product.

2. NSAID at the combination dose.

3. Codeine 60 mg (a lower dose should not be used even if the combination uses a lower dose of codeine, unless it is run as an additional group).

4. Aspirin (or acetaminophen) 650 mg.

5. Aspirin (or acetaminophen) 650 mg with codeine 60 mg.

The positive control combination (5) is already approved for "moderately severe" pain. Codeine 60 mg should be utilized (3) as a "least effective control group," and all other treatments in these studies should be superior to codeine. Both combination treatments should be superior to each of their individual constituents. The study combination should be equal to the aspirin (acetaminophen)/codeine combination product (q-value of 0.5 to 0.75).

C. Studies using potent parenteral analgesics (such as morphine) as control drugs should utilize a pain model(s) that has(have)the sensitivity to demonstrate a difference between two doses of the potent parenteral agent and a difference between the two components of the combination and the combination. Two studies in the same pain model or one study in each of two models should be performed with the following treatment groups:

1. Combination product.

2. NSAID at the combination dose.

3. Codeine 60 mg (a lower dose should not be used even if the combination uses a lower dose of codeine, unless it is run as an additional group).

4. Two different doses of the parenteral analgesic.

As above, it is anticipated that all other treatments will be superior to codeine 60 mg. The two doses of parenteral analgesic should be distinguishable. The combination should be comparable to one of the doses of the parenteral (q-value of 0.5 to 0.75).

These studies should include an assessment of the effects of the drug on the central nervous system (i.e., mood changes, drowsiness, etc.) as well as classic pain assessment.

Comparing effectiveness of drugs with different pharmacokinetic or pharmacologic effects may be difficult. In comparisons with approved agents in B. and C., in order to make claims of comparability or superiority to certain doses of the control drug, it will be necessary that the time-effect curves show comparability or superiority at all time points during the treatment interval. If the shape of the time-effect curves are different, e.g., if one drug peaks earlier but its effects have dissipated by 4 hours (common with narcotics) whereas the other drug peaks slower and lower but lasts 6 to 12 hours (common with NSAIDs), potency comparisons based on SPIDs (Sum of Pain Intensity Differences) and TOTPARs (Total Area Under the Pain Relief Curve) are not valid and comparability statements must be based on other measures or may not be appropriate at all.

Multiple-Dose Clinical Studies

In FDA's experience, single dose studies underestimate the incidence of adverse effects that will be seen in ordinary use of the drugs, which usually involves multiple doses. Similarly, adverse effects seen in trials of patients who require analgesics less than 2 weeks may not adequately reflect the incidence of adverse effects in patients who require pain medication more chronically.

A. In order to demonstrate efficacy in patients with usual short-term use and to accumulate safety data adequate to define whether the frequency of adverse effects associated with usual usage at the 1 to 3% level is grossly different from alternative treatments (i.e., more than a fivefold increase), there should be controlled studies extending 2 or more days involving multiple doses and measuring both safety and efficacy. These studies should be controlled by using a currently marketed standard drug appropriate for the patient population and should involve at least 100 patients on combination and at least an equal number of control patients. It is recognized that pain assessment in such studies is not as rigorous, as a rule, as in the single dose trials.

B. In addition, in order to define whether the adverse effects associated with chronic usage are grossly different from those observed with alternative treatments (see above), there should be controlled studies of at least 1 month's duration of daily dosing, measuring safety and efficacy, and using as a control a currently marketed standard drug appropriate for the patient population. These studies should involve patient groups of the same size as above, i.e., 100 patients per treatment group.

GUIDELINES FOR THE CLINICAL EVALUATION

OF DISEASE MODIFYING ANTIRHEUMATIC DRUGS (DMARDs)

"General Considerations for the Clinical Evaluation of Drugs" and the preceding guidelines for the evaluation of nonsteroidal anti-inflammatory drugs (NSAIDs) contain important information relevant to drug evaluations, and the suggestions contained in them are entirely appropriate to the evaluation of disease modifying or slow acting antirheumatic drugs (DMARDs or SAARDs). The term DMARDs has been used in these guidelines after much debate among expert rheumatologists. It is clear, however, that rheumatologists who prefer to call these drugs DMARDs rather than SAARDs recognize that disease modification in its literal meaning has only been demonstrated to a limited extent in limited studies for any drugs of this class to date. Those rheumatologists favoring the term DMARD are influenced by the fact that the SAARDs may show changes in some of the secondary manifestations of rheumatic diseases such as sedimentation rate, rheumatoid factor, cellular immune function, etc., and these guidelines encourage following these parameters when studying DMARDs. Certainly disease modification in the sense of "curing" the disease is beyond the promise of any of the currently available drugs and the use of the term DMARD in these guidelines should not be construed to suggest either to health professionals or to the lay public that these drugs have proven curative properties.

This guideline is intended to describe additional approaches to those mentioned earlier. Examples are given in order to provide guidance and are not meant to limit the manner in which the data for an NDA are developed.

I. INTRODUCTION

Disease modifying antirheumatic drugs are defined as agents thought to alter the stimulus causing the inflammation in rheumatic diseases rather than the inflammatory response itself. Typically, they are :

(1) Agents whose effects do not appear for weeks or months in contrast to the appearance of effects within hours to days with the more rapidly acting drugs, such as NSAIDs or corticosteroids.

(2) Agents that can be withdrawn without associated clinically detectable deterioration for a period of weeks after improvement has been obtained, in contrast to the flare of signs and symptoms that occurs within days of withdrawing NSAIDs or corticoids.

Typical DMARDs already approved for rheumatoid arthritis include antimalarials, parenteral gold salts, oral gold, D-penicillamine, and azathioprine. Cyclophosphamide and methotrexate have been recommended for approval by the Arthritis Advisory Committee. The modes of action of these agents in rheumatoid arthritis are currently unknown, although it is hypothesized that they may act by affecting the immune system. Some of them may be oncogenic.

Although DMARDs may be effective in other rheumatologic conditions, these guidelines refer only to rheumatoid arthritis as the prototypic disease in which DMARDs are to be used.

II. GUIDELINES FOR CLINICAL STUDIES OF DMARDs

General

Many of these drugs have potential toxicity (teratogenicity, oncogenicity, and/or effects on the immune system) that precludes their initial introduction into normal volunteers or into patients who are not yet candidates for currently available DMARDs. In addition, these drugs will usually need to be administered over many months to establish safety and efficacy. Therefore, initial planning should generally design trials that overlap Phase I through Phase III studies providing opportunities for patients willing to accept the risk of participating in an early study to be included in later studies in order to see if they can derive some benefit from the new drug. Nevertheless, for the sake of clarity, the phases will be separated in these guidelines.

A. Phase I Studies

See "General Considerations for the Clinical Evaluation of Drugs". Phase I studies are intended to determine levels of tolerance, early dose-ranging for safety and efficacy, and metabolism. As noted above, in most cases these drugs are not suitable for use in normal volunteers or patients with mild disease. Metabolism and pharmacokinetic data should be obtained as part of studies intended to evaluate effectiveness and safety. Similarly, even studies primarily intended to define tolerance should include evaluations of effectiveness.

1. Setting and Investigators

Phase I studies should be carried out by institutions with a full range of clinical and laboratory facilities and the patients should be kept

under close observation. Further, the studies should be under the direction of experienced clinical pharmacologists and/or rheumatologists as well as experts in other appropriate areas of investigation suggested by preclinical and in vitro testing which shows potential effects on immune or progenitor function.

2. Subjects

If the preclinical studies show no mutagenic, teratogenic, immune system effects, or other serious effects at or near the expected therapeutic range, Phase I studies can be in normal volunteers. If, on the other hand, significant effects of these types have been demonstrated or might be possible, e.g., a derivative of a drug known to have such effects or which show interference or alteration of immune (e.g., effects on T or B cells) or host defense mechanisms (e.g., effects on macrophages or leukocyte), careful selection of an appropriate patient population is needed.

If rheumatoid arthritis patients are used, they should have definite or classical rheumatoid arthritis by ARA criteria. Further, these patients should be without other serious medical conditions. Patients who ordinarily would be prescribed a DMARD would be appropriate for Phase I studies of a new or potential DMARD, if the preclinical or in vitro studies did not show mutagenicity and if, even though there might be immune effects, animal toxicity studies showed no toxicity due to those effects. The criteria used to select these patients should generally be more restrictive than those used in later studies to ensure that the patients have sufficiently severe disease to justify the risk of a relatively unknown agent and that they do not have other medical problems that may increase the risks. Patients should also have disease in which change due to therapy can be measured. Patients who have failed to respond to a standard DMARD or have had side effects requiring their discontinuation should be considered as candidates for drugs that animal or in vitro studies suggest may be potentially more dangerous. The guiding principle should be that if there is potentially greater risk associated with the new agent, patients should have more serious or more recalcitrant disease. In any case it is particularly important that informed consent be complete and some provisions be made to assess that patients understand what they are consenting to.

When long-term or short-term use of a drug may affect gonadal function, as indicated by preclinical studies or previous use, women of non-childbearing potential, or women and men not wishing to parent children should be chosen for Phase I studies.

Corticosteroids (up to 10 mg prednisone equivalent daily) and/or NSAIDs may be continued in patients with rheumatoid arthritis used for Phase I studies of DMARDs.

3. Observations

a) Toxicity

Examinations for toxicity should include a careful search for unwanted drug effects, paying particular attention to organ systems (including the immune system) demonstrated to be affected by the DMARD in preclinical and earlier clinical studies. Dose-extending and duration-extending studies primarily undertaken to establish safety should proceed and provide a safety margin for studies that are primarily concerned with evaluating efficacy and have less intensive laboratory monitoring than the studies that are primarily for toxicity testing. In testing for toxicity, cognizance should be taken of experience with related drugs, of preclinical experience in animal studies with the DMARD under investigation, and of the general characteristics of the DMARD. Some DMARDs are known to have delayed toxicity, e.g., delayed ocular toxicity with antimalarials; delayed hepatic fibrosis with methotrexate, delayed oncogenicity with cyclophosphamide. Therefore, a prolonged period of observation after drug exposure is appropriate for all subjects exposed to the drug.

b) Pilot Efficacy Observations

Based on the general experience with known DMARDs, it is unlikely that efficacy can be ruled out in less than 6 months although it may be seen in shorter periods. Initial assessment of effectiveness should be made as part of the exploratory toxicity studies that, with more extensive monitoring, establish or exceed the limits of dose and duration for the efficacy studies.

c) Metabolic Studies

When patients have participated in early metabolic and kinetic studies and are continued on the DMARD, a repeat study after the drug has exhibited its effects (e.g., 3 to 6 months later) may be desirable to document any changes in metabolism and kinetics consequent to any changes in disease activity (e.g., the effect of the subsidence of liver disease on the drug's metabolism). Reliable data from studies in other countries (if adequate and available for review) may be used to provide guidance and are acceptable for part of the Phase I studies.

B. Phase II Studies

Early Phase II studies should extend the observations on dose related toxicity and should establish the effectiveness of the drug. Because of the nature of the DMARD class, most of the Phase II studies will be of at least 3 to 6 months duration. Because the response to DMARDs is so delayed, titration to effectiveness designs are not feasible. It is therefore recommended that the earliest studies be blinded, parallel, dose-response studies so that the larger Phase III trials can be focused on the right dosage range. The earlier trials should cover a relatively wider range and ideally should include, if possible, a placebo group so that a failure to show a dose response can be properly interpreted, i.e., all doses are effective versus all doses are ineffective.

During early Phase II studies patients should be under close observation. The number of subjects needed to document the appropriate dosage range will vary depending on the drug's efficacy. It is not advisable to proceed into Phase III studies until there is good assurance that the drug is being used at the appropriate dose.

C. Phase III Studies

General: Phase III studies consist of controlled clinical trials designed to demonstrate effectiveness, further refine dosage, and better define the nature and incidence of the most common adverse reactions. Phase III studies are substantially similar to Phase II controlled trials except that larger numbers of investigators and patients become involved in both expanded controlled and uncontrolled trials and the dose ranges used should be narrower.

1. Setting

Most of these trials take place in a controlled outpatient setting, although some may be started in a hospital. Broad spectrum hospital and medical services must be readily available, if needed. If a multiclinic trial system is used, it is preferable that each contributing clinic evaluates a sufficient number of patients to allow statistical analysis for investigator by treatment interaction (usually more than 12 patients per center).

2. Investigators

Phase II studies should be under the direction of experienced rheumatologists with experience in the testing of drugs and long-term

management of patients with rheumatoid arthritis. In Phase III, rheumatologists or internists with rheumatologic experience, familiar with the use of DMARDs, are appropriate.

3. Subjects

All participants should have active definitive or classical adult rheumatoid arthritis, as defined by the ARA criteria. Most patients should be functional Class II or Class III, and should have disease in which change due to therapy can be measured. It may be useful to measure the anatomical stage of the disease or other possible predictors of response or toxicity in order to define populations of patients likely to respond to the DMARD, or likely to suffer with toxicity from its use.

Generally, a patient entering a trial of a DMARD should have had an adequate and unsuccessful trial of drugs considered to entail less risk. The criteria used to select and exclude patients should be very specific. If there is no previous experience in man or if animal toxicity shows mutagenicity, teratogenicity, or tumorigenicity, then patient selection criteria for Phase II and early Phase III generally should include failure to tolerate or respond to an adequate trial of at least one DMARD. Most patients would presumably be receiving NSAID and/or steroids upon entry. These may be continued during the trials, but rules for the management and recording of concomitant medication should be part of the trial protocol.

Women of childbearing potential, nursing mothers, and men wishing to parent children should not be chosen for testing until animal reproductive and mutagenicity testing is completed. Positive animal findings are not a contraindication to participation of patients who are at risk based on the animal findings, however, the drug should be a "last resort" and patients must be explicitly informed of the animal findings.

4. Observations

Efficacy can be demonstrated in three areas: first, reduction of the signs and symptoms of inflammation*, second, retardation of structural damage, and third, induction of disease remission.

Although substantial information on the efficacy of DMARDs may be gained with a trial of 6 months' duration, once marketed, these drugs may be administered to patients for many years. Therefore, the majority of the trials in an NDA or those reported in the

---

* These parameters are identical to those used to evaluate anti-inflammatory agents (see earlier section of these guidelines).

literature that last 6 to 12 months are considered minimal. There should be some additional studies designed to continue treatment in patients who demonstrate continued need for the drug, in addition to studies that test in a well-controlled trial the benefits and risks of continuing versus discontinuing long-term treatment.

Definitions of the degree of response (e.g., excellent, good, fair, or poor) should be specified in the protocol. Further specific criteria for withdrawal from the study, for toxicity, and for lack of efficacy should also be part of the protocol.

Rheumatoid arthritis is a disease characterized by a variable course. Before actually starting a trial it is desirable to obtain several baseline measurements over a short stabilization period (e.g., 1 month). During this stabilization period, the patient should be taking adequate, standard background therapy (NSAID and/or corticosteroids) for rheumatoid arthritis. As functional capacity and several of the measures of disease activity respond to brief bursts of NSAID or steroid administration, it is most desirable that all evaluations (but especially those at the beginning and end of the trial) be carried out with background drug therapy as stable as possible. On the other hand, a reduction in background drug therapy may be considered a primary efficacy parameter; use of this as a parameter should be reflected in study design rather than being included after the fact. Generally, the measures of disease activity should be recorded at the beginning of the trial and every 3 months thereafter unless animal and/or earlier human studies suggest the drug exerts a more rapid response than currently available DMARDs. In this case, more frequent or differently timed assessments might be indicated.

At the beginning and the end of the trial and at appropriate intervals in between, serological (including quantitative rheumatoid factor or similar quantitative tests) assessments should be undertaken. Although these tests are not considered efficacy variables it will be useful for labeling to be able to report the effect, if any, of the drug on any tests that are commonly measured in these patients. It may be valuable to obtain and store serum samples at regular intervals so that sera are available for future serial immunological assays of potential interest. Follow appropriate storage procedures!

Although radiological progression of disease or healing of erosions are interesting and potentially useful measures of efficacy, there are inadequate data, at present, on the optimal interval to assess x-ray progression of disease. There are also methodological problems in the standardization of taking and reading the films. Most investigators feel that at least a 1-year interval (and preferably 18 to 24 months) is needed to examine treatment

effects on radiographic progression. Films should include at least the hands and wrists but other joints may be added to the series if desired. When doing radiological examinations, the following sources of variation must be addressed: (1) inter- and/or intra-observer variability; (2) the background and experience of the readers; (3) the radiologic techniques; (4) the disease duration and stage; (5) the use of a standardized scoring system. A serious limitation to the use of radiographic analysis is matching the comparison or control group for factors which may affect disease progression.

Evaluations for adverse effects will usually need to be done more frequently than efficacy evaluations, at least initially. The type and frequency of laboratory and clinical screens needed to detect unwanted effects will vary substantially from one drug to another, depending on data from preclinical animal studies, Phase I information, and knowledge of drug effects from previous use for other indications. Attempts should be made to standardize the observation and reporting of side effects as part of the protocol. More frequent observations, both clinical and laboratory, should usually be carried out in the first group of patients (e.g., weekly). As information about adverse effects accumulates, some measurement intervals may be extended, for example, to biweekly, monthly, and then to every 3 months. In general, adverse reaction screening should be done at least every 3 months during the IND/NDA study period. For chronically administered drugs, special studies for chronic toxicity, such as ophthalmic examination, should be done early, and in enough patients to provide an adequate number of patients with 2 years exposure at the time of NDA submission. Observations for toxicity by "blinded" observers as well are desirable. Because patients with rheumatoid arthritis treated with DMARDs are usually on NSAIDs, frequently on corticoids, and are also subject to complications from their disease, it is recommended that positive control comparisons with alternative approved DMARD treatment be included in Phase III.

5. Clinical Designs

At this stage of our knowledge, crossover trials for DMARDs are not recommended because treatment periods with DMARDs are so long and potential carry-over effects may occur. In most cases, multicenter trials are necessary to enlist sufficient numbers of patients for statistical analysis.

Physicians or other qualified observers evaluating efficacy and toxicity of a DMARD should be "blinded" with respect to the patient's treatment. In cases where blinding is difficult because the drug requires close monitoring for effects that are used to adjust dosage, one physician can be assigned to care for the patient, with access to lab. data, etc., and a second physician, who

is not privy to these data, can evaluate the patient periodically to determine efficacy and clinical toxicity of drugs (this second investigator should not be involved in the patient's care in any way and steps should be taken to isolate him/her from day-to-day contact with the physician and other health personnel monitoring treatment).

In order to develop the body of information necessary for approval of a DMARD, studies using the following different control groups should generally be conducted:

(1) Comparison of the drug with a placebo -- a double blind trial in which a DMARD is compared to placebo in parallel groups of patients. Such trials are usually not longer than 6 to 9 months and are followed by an open phase during which all patients take the DMARD. Continuation of the DMARD in all patients who complete the trials provides an inducement to both physicians and patients to participate in a placebo controlled trial because all patients will eventually have an opportunity to take the new drug. If too many patients discontinue treatment because of lack of efficacy, however, it jeopardizes the ability to demonstrate efficacy at 6 to 9 months. In addition, if there are no control patients for comparison of adverse effects, it is difficult to separate those adverse effects due to the drug from those associated with the disease or other drugs.

(2) Double-blind comparisons of different doses of the DMARD -- it is important as part of the IND/NDA process to collect dose-response data. It may be difficult, however, to study more than 2 doses of a DMARD in one study. For this reason, it may be necessary to use several studies comparing different pairs of doses, e.g., one study studying x and 3x another x and 1/3x. The doses within each study, as in the example, should generally be different by a factor of threefold or fourfold. Historically, the dose of DMARDs has "drifted" downward with continued clinical experience. The dose being compared against placebo or against another DMARD, presumably the estimated best dose after Phase II, should also be compared to a lower dose in a long-term study. If the drug's toxicity does not preclude it, a similar study using a higher dose is also recommended. The number of patients per group in these trials will generally need to be greater than in placebo controlled trials, because the differences between two active treatments is usually smaller than the difference between any effective dose and no treatment (placebo).

(3) The new DMARD, compared to an already accepted standard drug, such as gold, penicillamine, azathioprine, etc. -- in a randomized double-blind parallel controlled trial. The selection of the appropriate dose of the standard drug is critical, particularly if there is a large range of recommended doses. If showing the similarity of an effect of the new agent and a standard agent is critical to establishing effectiveness, again larger numbers of patients per group are generally required. Interpretation of an active control trial requires assessment of the ability of the trial to have shown a difference between treatments (see the discussion of "q" in item A. 5. under the Clinical Studies section of the Rheumatoid Arthritis Guidelines for Nonsteroidal Anti-inflammatory Drugs).

The number of patients to be entered into a DMARD trial depends upon the effectiveness of the agent, the magnitude of the difference in disease activity one wants to detect, and the number of patients who drop out due to adverse effects or lack of efficacy. The importance of good estimates of these three factors prior to the trial cannot be over emphasized. In general, analyses should be planned to include as well as exclude dropouts, and there should be a discussion of why the analyses show differences if they do.

If it is anticipated that the new DMARD will be similar to those currently available and will probably be used chronically, then prospective, carefully collected data should be available in at least 400 patients for 1 year and 200 patients for over 2 years at the time of NDA approval. These may be patients who have participated in previous controlled studies or who have been entered into open studies.

Potentially useful drugs with a good safety record in the later stages of Phase II/III may be considered for study in children with juvenile arthritis.

III. Recommendation for Registry

During the IND/NDA process, common adverse effects (those with an incidence above 1%) occurring during the first year of treatment are usually identified. Unusual events (those occurring in less than 1 in 100 patients) and adverse events associated with long-term therapy (greater than 1 year) have not been seen or at least have not been associated with the drug. There is usually limited experience with these drugs in elderly patients, patients who have multiple diseases, or in patients who are on multiple drugs for other diseases unrelated to their rheumatic disease. To minimize the risk to subsets of patients not heavily represented

in the clinical trials, and at the same time obtain optimal benefit from the introduction of new drugs to the majority of patients as soon as possible, it is essential to establish plans for long-term surveillance of some patients.

For this reason, it is recommended that patients in study protocols of DMARDs be informed of the need for possible long-term follow-up, and that a list of these patients' names, status, and location be maintained by the study sponsors after marketing. These patients could serve as a "leading edge" of longer term exposure.

REFERENCES

1. Ropes, M.W., G.A. Bennet, S. Cobb, R. Jacox, and R.A. Jessar, "1958 Revision of Diagnostic Criteria for Rheumatoid Arthritis," Arth. Rheum., 2:16-20, 1959.

2. Cooperating Clinics Committee of American Rheumatism Association, "A Seven-Day Variability Study of 499 Patients with Peripheral Rheumatoid Arthritis," Arth. Rheum., 8:302-334, 1965.

3. Steinbrocker, O., C.H. Traeger, and R.C. Batterman, "Therapeutic Criteria in Rheumatoid Arthritis," JAMA, 140:622-659, 1949.

4. Hess, E.V., "A Standard Database for Rheumatic Diseases," Arth. Rheum., 17:327-336, 1974.

5. American Rheumatism Association, JRA Criteria Committee; "Proposed Revision of JRA Criteria," Arth. Rheum., Supplement on Childhood Rheumatic Diseases, January 1977.

*U. S. Government Printing Office 1988: 201-874/83634