Guidance for Industry

Systemic Lupus Erythematosus — Developing Drugs for Treatment

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact Joel Schiffenbauer (CDER) 301-827-2090 or Jeffrey Siegel (CDER) 301-827-5096.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

March 2005
Clinical/Medical

Guidance for Industry

Systemic Lupus Erythematosus — Developing Drugs for Treatment

Additional copies are available from:

Office of Training and Communications
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

March 2005
Clinical/Medical

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 1

III. MEASUREMENT OF DISEASE ACTIVITY AND CLINICAL OUTCOMES 2

A. Disease Activity Indices 2

B. Flares 3

C. Damage 4

D. Organ-Specific Indices 4

E. Health-Related Quality of Life and Fatigue 6

F. Serologies 6

IV. SLE CLAIMS 6

A. Reduction in Disease Activity of SLE 6

B. Effectiveness in the Treatment of a Specific Organ System Manifestation 7

C. Complete Clinical Response/Remission 10

D. Reduction in Flares 10

V. TRIAL DESIGN AND ANALYSIS 10

A. Phase 2 Trials 11

B. Efficacy Trials 11

1. Disease Activity Trials 11

2. Lupus Nephritis Trials 13

3. Other Organ-Specific Claims 13

C. Studies to Show Superior Safety 14

D. Other Trial Design Issues 14

1. Concomitant Medications 14

2. Issues of Blinding 14

3. Standard of Care Issues 15

4. Extension Trials 15

5. Trial Duration 15

VI. SURROGATE MARKERS AS ENDPOINTS 15

VII. RISK-BENEFIT ASSESSMENT 16

VIII. LUPUS AND PHARMACOKINETICS 17

A. General 17

B. Special Studies 17

C. Drug Interactions 17

REFERENCES 18

APPENDIX: GLOSSARY OF ACRONYMS 19

Guidance for Industry[1] Systemic Lupus Erythematosus — Developing Drugs for Treatment

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION

This document is intended to provide guidance to industry on developing drugs for the treatment of systemic lupus erythematosus (SLE). The following topics are covered:

- Outcomes and measurements of lupus disease activity, including the use of disease activity indices, flares, and organ-specific outcomes
- Indications that the Agency may be willing to approve for new drug therapies for lupus
- General trial design issues, the use of surrogate endpoints in relation to lupus, and the overall risk-benefit assessment that needs to be addressed for any new therapy of lupus
- Issues related to lupus and pharmacokinetics

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

Systemic lupus erythematosus is a chronic disease characterized by protean manifestations often demonstrating a waxing and waning course. Whereas in the past a diagnosis of SLE often implied a decreased life span due to internal organ system involvement or to toxic effects of therapy, recent improvements in care have dramatically enhanced the survival of SLE patients with the most severe and life-threatening manifestations. Unfortunately, current treatments for

[1] This guidance has been prepared by the Division of Anti-Inflammatory Analgesic and Ophthalmologic Drug Products in the Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration.

SLE remain inadequate as many patients have incompletely controlled disease, progression to end-stage organ involvement continues, and current therapies carry potential risks of debilitating side effects. Therefore, it is important to clearly describe acceptable study endpoints to establish efficacy to facilitate the development of novel therapeutic agents which have the potential to be more effective and/or less toxic.

Although many patients with SLE exhibit symptoms that involve the skin and joints, other symptoms of SLE vary widely among patients. No single biological mechanism explains the varied manifestations of disease. Disease activity scores allow a comparison of disease severity in SLE patients whose disease affects different organ systems. Several such indices reliably measure disease activity in SLE patients in varied settings. Some of these indices mirror the assessment of experienced clinicians and are sensitive to changes in disease activity. One of the scoring systems, the British Isles Lupus Assessment Group (BILAG), scores patients based on the need for alterations or intensification of therapy. Thus, these indices can be used as endpoints to establish efficacy.

It is uncertain whether the SLE disease activity indices will clearly delineate important clinical responses to therapy in all situations. Some treatments may target a biologic mechanism which selectively underpins only certain lupus manifestations, or only those related to a single organ system. In these situations, an organ-specific measure of disease activity may be a preferable outcome measure. This guidance addresses claims of improvement in overall activity of SLE, as well as claims of improvement in organ-specific manifestations of SLE such as lupus nephritis. It is important that any therapy that claims to improve disease in one organ system not worsen disease elsewhere. In addition to the primary outcome measure selected for a given trial in SLE, every trial should also assess other aspects of the disease process, as this information may be informative about the overall risk-benefit assessment (see Section VII, Risk-Benefit Assessment).

This guidance document first provides a general discussion of outcomes and measurements of lupus disease activity including the use of disease activity indices, flares, and organ-specific outcomes. The document then presents the claims that the Agency may be willing to approve for new drug therapies for lupus. Following this, the document presents general trial design issues, discusses the use of surrogate endpoints in relation to lupus, the overall risk-benefit assessment that needs to be addressed for any new therapy of lupus, and, finally, briefly presents some issues related to lupus and pharmacokinetics.

III. MEASUREMENT OF DISEASE ACTIVITY AND CLINICAL OUTCOMES

A. Disease Activity Indices

The clinical measurement of disease activity in SLE involves an assessment of the characteristic signs and symptoms of disease and the results of laboratory parameters. Academic and clinical investigators have identified those measures they believe are important for evaluation in clinical trials. These parameters include a measure of disease activity, a measure of disease-induced

damage, a measure of therapy-induced damage, a measure of response as determined by the patient (i.e., *a patient global response*), and a measure of health-related quality of life (HRQL).

Although patterns of stable, increasing, or decreasing disease activity form the basis for initiating or adjusting treatment in SLE, the specific manifestations that characterize the level of disease activity vary considerably from patient to patient and at different points in time. Indices of disease activity have been developed that correlate with assessments of panels of expert clinicians. These indices score disease manifestations using predefined criteria based on the presence or absence of different aspects of the disease or, in the case of the BILAG, on the clinician's assessment of the need to change therapy. In clinical studies, these indices have been shown to be valid based on the concordance of scores with expert opinion, acceptable interobserver variability among trained evaluators, correlation between individual patients' scores on different indices, and correlation between increases in scores and clinical decisions to increase therapy. The SLE Disease Activity Index (SLEDAI and SELENA-SLEDAI), the BILAG, the SLE Activity Measure (SLAM), and the European Consensus Lupus Activity Measure (ECLAM) have been shown in cohort studies to be sensitive to change in disease activity (Strand 1999) and can be used in clinical trials. It is important that analyses of disease activity measures be defined prospectively, and they can include comparisons of change in disease activity scores or in disease activity. We recommend prespecifying in the protocol statistical approaches regarding, for example, dropouts or missing data.

There has been considerable interest in the development of a responder index to measure response to therapy on an individual basis. Some proposed definitions of a responder specify a minimum improvement in a measure of disease activity with no worsening in other aspects of lupus. A responder index would allow a clinical trial to determine directly what proportion of patients had a clinically meaningful improvement from therapy. It is important that such a responder index be assessed for reliability, face validity, content validity, and sensitivity to change to be fully validated. Full validation would also include a demonstration of the ability to discriminate treatment with a known active agent compared to an inactive control in a clinical trial. Exploring the use of responder indices in prospective studies will help determine the utility of these measures in clinical trials. At present, there are no generally accepted and validated responder indices in lupus.

B. Flares

The clinical course of SLE is generally characterized by periods of relatively stable disease followed by flares of disease activity. Studies that measure disease activity at fixed time points may miss flares in between study assessments. In one study, rates of flare were measured at an average of 0.6 flares per year (Petri 1991). A *flare* should reflect an episode of increased disease activity and should correlate with a need for increase in or change in treatment on clinical grounds. Criteria for major flare might include initiation of high dose glucocorticoid therapy, a change in dose of immunosuppressive therapy, hospitalization, or death. The frequency of flares may be affected by gender, menopausal status, treatment, and other patient characteristics. We recommend prospectively defining *flare*.

C. Damage

Patients suffering from lupus experience irreversible damage to internal organ systems. Accumulation of damage occurs over a period of years. Therapy-induced organ damage may also occur. An index of organ damage was proposed and validated as the Systemic Lupus Erythematosus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index. Validation studies show that high scores on the SLICC/ACR Damage Index are predictive of increased mortality, and damage in the renal and pulmonary components are associated with poor outcomes (Stoll 1996). The prognostic information derived from SLICC/ACR Damage Index scores suggests they may be useful as stratification variables for clinical trials. The SLICC/ACR Damage Index measures only changes that have been present for at least six months; therefore, only longer-term clinical trials could demonstrate reduction in the rate of progression of damage using this measure. Some of the components of the SLICC/ACR Damage Index are measures of toxicity related to current treatment modalities. Use of the SLICC/ACR Damage Index as outcome measures in clinical trials could be complicated if a new therapy were associated with toxicities not measured by the Damage Index, or if the use of organ damaging concomitant treatments were not balanced between the groups. The SLICC/ACR Damage Index can be used as an endpoint, but we recommend discussing this with the appropriate reviewing division before beginning trials.

D. Organ-Specific Indices

Organ-specific measures of disease provide another approach to assessing disease activity in lupus. To measure organ-specific disease activity in a clinical trial, a responder analysis could be applied by measuring if subjects demonstrate improvement in the involved organ system using prespecified criteria, such as components of validated disease activity indices if these components can be shown to reflect disease activity. Examples of issues related to studies of renal and skin involvement are provided below. We recommend investigators propose outcome measures for specific organs studied.

Lupus nephritis is the most commonly studied organ-specific manifestation of lupus. The presence of diffuse proliferative (WHO class IV) and severe focal proliferative (WHO class III) glomerulonephritis in patients with SLE who have measures of inflammatory activity and damage is associated with increased long-term risk of progression to end-stage renal disease and mortality. Patients with severe lupus nephritis are often treated with high doses of immunosuppressive agents, including cyclophosphamide, and high doses of corticosteroids. These regimens are based on studies that suggest a decrease in the long-term risk of progression to end-stage renal disease. The outcome of lupus nephritis has improved markedly in recent years with 5-year survival rates of 90 percent or greater and 10-year survival rates of more than 80 percent reported (Urowitz 1999). However, there remains a need for additional regimens as current treatments can be highly toxic and not effective in all subjects.

After a diagnosis of lupus nephritis is established, disease activity is assessed clinically by examination of the urinary sediment and by measures of renal function. A variety of outcome measures have been used in clinical trials of lupus nephritis to assess organ-specific disease

activity. Mortality is an important outcome measure, but low mortality rates and long observation times make it a relatively insensitive measure in clinical trials. Measures of renal function can be used as outcome measures, including progression to end-stage renal disease (ESRD), sustained doubling of serum creatinine, creatinine clearance, and iothalamate clearance, for full approval. Other measures may also be suitable and can be employed in therapeutic studies if sufficient data to support the proposed measure are available. The use of the doubling of serum creatinine is the best-validated of these measures as it has been shown to reliably predict long-term renal outcomes; however, it is insensitive to smaller changes that represent earlier signs of damage that are nonetheless clinically important. Changes in the urine protein/creatinine ratio may serve as an indicator of the need for further assessment with a 24-hour urine collection for quantitation of the extent of proteinuria and impairment in renal function as measured by creatinine clearance. We recommend investigators design trials to minimize confounding variables (Boumpas 1998) as these can complicate interpretation of renal function measures, including serum creatinine and creatinine clearance.

Changes in urinalysis can provide important information for the assessment of renal inflammation in lupus nephritis. The presence of cellular casts and hematuria, when measured accurately, is considered a sensitive indicator of the level of activity of lupus nephritis. However, central laboratories may be unreliable in assessing the presence of casts as they can break up during transport. There is no consensus on the appropriate evaluation of urine sediment. Local or central laboratories could be used if the chosen method is shown to be accurate and reproducible.

Major flares of lupus nephritis, as assessed by urinary sediment, proteinuria and renal function, have been used as outcome measures in clinical trials. Patients who experience nephritic flares characterized by nephritic sediment and an increase in serum creatinine or decrease in glomerular filtration rate (GFR) may be at increased risk of developing a persistent doubling of serum creatinine. Renal remission in response to therapy has been defined as a return to normal levels of an elevated creatinine and proteinuria and normalization of nephritic sediment. Patients who fail to normalize an elevated serum creatinine in response to therapy may have an increased risk of progression to renal failure (Levey 1992). Assessment of proteinuria is particularly important in patients with membranous glomerulonephritis; however, this is a less common form of lupus nephritis. Increases in proteinuria in patients with other forms of glomerulonephritis may not translate into unfavorable long-term outcomes, and, therefore, measures of proteinuria are not adequate to address clinical outcomes.

Skin is one of the organs most involved in SLE. The most common of the skin manifestations include discoid lupus, malar rash, subacute cutaneous lupus, and alopecia. Photosensitivity and oral ulcers are additional common manifestations. A variety of outcome measures can be used in clinical trials to assess the efficacy of new therapies on skin disease including erythema, induration, scaling, and physician and patient global assessment. In addition, outcomes such as involved surface area changes and skin biopsies can be considered. Investigators can propose additional or alternative outcome measures depending on the type of skin disease studied. It is also important to differentiate irreversible damage from active disease, as it would not be amenable to therapy.

E. Health-Related Quality of Life and Fatigue

The Agency recommends that HRQL measures be studied in all trials of SLE. Instruments that assess health status and HRQL may measure aspects of SLE and its impact on patients that are not fully assessed by other outcome measures. It is important that trials showing improvement in a specific organ or in disease activity demonstrate no or minimal worsening in measures of HRQL. Patients with active SLE may have increased disability as assessed by the Health Assessment Questionnaire (HAQ) or Modified Health Assessment Questionnaire (MHAQ). Health-related quality of life has been assessed in lupus patients using a number of generic instruments including the HAQ, MHAQ, Arthritis Impact Measurement Scale (AIMS), the Medical Outcomes Survey Short Form-20 (SF-20), and Short Form-36 (SF-36). Differences compared to controls have been observed in several domains and subdomains. Some instruments do not adequately assess fatigue, an important symptom for many lupus patients. Specific instruments have been studied for assessment of fatigue (e.g., the Krupp Fatigue Severity Scale (KFSS)). As with any instrument, HRQL instruments used in clinical trials of SLE should undergo validation regarding content validity (inclusion of all relevant domains), construct validity, sensitivity to change, and other criteria. The use of these outcomes is critical to understanding both the efficacy of an agent as well as its potential adverse events. Even if the measure does not improve with a specific therapy, it should not worsen. Improvement in HRQL alone would not result in approval at this time.

F. Serologies

Serologic markers play an important role in the assessment of disease activity in SLE, including assessment of anti-double-stranded DNA, complement levels, and others. Serologic markers are critical for understanding the pathogenesis of disease. Serologic markers have an imperfect correlation with disease activity and cannot substitute for a direct assessment of clinical benefit. We recommend studying serologic marker data in clinical trials. These data, in conjunction with clinical measures, may play a role in assessing clinical outcomes and identifying potential clinical benefit from new therapies. Serologies can serve as supportive evidence of efficacy at this time (see Section VI, Surrogate Markers as Endpoints).

IV. SLE CLAIMS

We may be willing to approve the following claims for SLE if supported by substantial evidence: (1) reduction in disease activity; (2) treatment of lupus involving a specifically identified organ (e.g., lupus nephritis); (3) complete clinical response/remission; and (4) reduction in flares.

A. Reduction in Disease Activity of SLE

This claim is intended to reflect clinical benefit associated with reductions in the signs and symptoms of SLE disease activity. SLE is a disease of long duration, with a waxing and waning course; therefore, this claim would ordinarily be established by trials of at least 1 year in duration. For products that may elicit the formation of antibodies, it is important that the clinical trials assess whether antibodies are formed and if they adversely affect efficacy and safety. We

recommend using methods that assess the activity of disease over the duration of the study in conjunction with methods that measure disease activity at the beginning and end. As part of any trials in support of this claim, we also recommend studying measures of damage and HRQL, as well as determining a patient global assessment. A validated disease activity index (DAI) is an acceptable outcome measure to demonstrate a reduction in signs and symptoms of SLE.

In a randomized clinical trial, the SELENA-SLEDAI, the SLAM, the BILAG, the ECLAM, or other established index could be used to measure disease activity. To represent a clinical benefit, the change in DAI should be both statistically significant and clinically meaningful and prospectively defined. Since the BILAG evaluates patients based on the need for additional treatment, the clinical interpretation of a change in score is apparent. A success in a 1-year trial could be defined as a greater reduction in the BILAG score at 1 year along with supportive evidence of reduction in monthly measurements of the BILAG score compared to controls (see also Section V.B.1, Disease Activity Trials, for a discussion of landmark versus area under the curve (AUC) analyses). For other indices, deciding whether changes in score are clinically meaningful may be more complicated. If a disease activity measure other than the BILAG is chosen, confirmation of a positive result with two different DAIs would be important to confirm the findings.

B. Effectiveness in the Treatment of a Specific Organ System Manifestation

In general, appropriate outcome measures in organ-specific trials are defined by the specific organ under study. For each organ studied, these include: (1) stabilization (no worsening of disease activity in the designated organ); (2) partial response; (3) complete response but still receiving medications; (4) complete remission (no ongoing treatments); (5) flares (time to flare and/or number of flares); and (6) ability to taper concomitant corticosteroids by clinically significant amounts. If corticosteroid dose is chosen as the endpoint, we recommend addressing the use of flexible dosing versus forced tapering. We also recommend addressing in the analysis plan the potential need for rescue medication.

For products being proposed for use in the manner of a specified short course of treatment leading to induction of a sustained remission, studies of 3-6 months duration may be acceptable with longer term follow-up for safety and durability of response. For products being proposed for chronic use, studies as short as 1 year may be considered.

We recommend that trials to demonstrate effectiveness in the treatment of a specific organ also include measures of overall disease activity, damage, and HRQL. Ideally these measures should improve in a clinically meaningful fashion.

Claims using the organ-specific approach may be either for the treatment of each organ studied (e.g., lupus nephritis) or for the treatment of lupus, depending on the number of patients and the type of organ impairment studied. To obtain approval for such a claim, you should show that there would be no worsening in terms of a patient global assessment as well as health-related quality of life.

Trials intended to study clinical benefit for specific organ systems could enroll subjects with disease affecting a single organ system (e.g., lupus nephritis). Patients enrolled in studies evaluating multiple organ systems can be stratified according to the specific organ system involved for randomization and analysis. It is important that the definition of a response be prospectively specified for each organ system under study. Trials of patients with disease activity affecting specific organ systems can define success as an increase in the proportion of responders among patients receiving study drug compared to controls.

Trials designed to assess efficacy of a product for the treatment of lupus nephritis should demonstrate an improved outcome for patients with biopsy-proved severe glomerulonephritis (WHO grades III or IV), or membranous glomerulonephritis. Short-term benefits may not reliably predict long-term outcomes; therefore, trials of lupus nephritis should be at least 1 year in duration. The following outcome measures could establish efficacy in lupus nephritis:

1) *Incidence of mortality and progression to end-stage renal disease.* Mortality and ESRD (when clearly defined prospectively) are objective, reliably determined, and the endpoints of ultimate importance. However, studies using these as the endpoint will generally require longer duration and larger sample size than may be needed when other endpoints are used.

2) *Sustained doubling in serum creatinine or other measure that has been validated including approximations of GFR such as iothalamate clearance or creatinine clearance studies*.[2] Doubling of serum creatinine has been shown to be associated with progression to ESRD. Thus, a decrease in the proportion of subjects meeting this endpoint in the treatment group compared to controls can be interpreted as demonstrating a patient benefit. Lesser degrees of change or changes in other measures may be considered but should be further justified. Similarly a significant change in GFR which has clinical importance may be considered. We recommend that sponsors provide data to demonstrate that these changes or other proposed measures are associated with a true clinical benefit (e.g., a significant reduction in the rate of progression to ESRD).

   A success in a trial utilizing this outcome measure would be defined as a decrease in the proportion of subjects whose serum creatinine attains a level double that of the baseline value and remains doubled for at least six months. Alternatively, a success in a trial could be defined as a reduction in the proportion of subjects experiencing a sustained fall in GFR of 50 percent or more.

3) *An unvalidated surrogate marker for lupus nephritis reasonably likely to predict clinical benefit.* FDA regulations for accelerated approval of new therapeutic agents (21 CFR 314, subpart H and 21 CFR 601, subpart E) provide an additional framework for FDA approval of drugs intended to treat serious or life-threatening diseases. One approach is to base approval on the effect on a surrogate marker, provided that specific criteria are met, and there is a commitment to verify the actual clinical benefit of the agent in studies completed after approval. Demonstration of marked and sustained improvement in renal function and renal inflammation in a seriously affected population of patients with lupus

[2] S*urrogate for development of ESRD; see Section VI on use of surrogate endpoints.*

glomerulonephritis may qualify for consideration under these regulations. Data showing that the measure of improvement is associated with improved patient outcomes can contribute to supporting the conclusion that the surrogate is *reasonably likely* to predict clinical benefit. Sponsors are urged to consult with the relevant FDA staff before embarking on a clinical program based on these regulations.

Use of the accelerated approval pathway for a product for lupus nephritis, for example, would necessitate the timely completion of studies of long-term clinical outcomes postmarketing. The verification of clinical benefit can be a difficult task. It is important that the necessary studies be a clearly described part of the clinical development program at the time the studies of the surrogate endpoint are undertaken.

4) *Induction of renal remission.* Active lupus nephritis is associated with evidence of renal inflammation, including cellular casts, proteinuria, and decreases in renal function. Organ-threatening WHO class III and IV lupus nephritis is frequently treated with cyclophosphamide and high doses of corticosteroids, agents that are associated with significant toxicity. A treatment that induces a sustained remission in lupus nephritis would confer a clinical benefit. Clinical studies of lupus nephritis use varied definitions of renal remission, but generally specify decreases in hematuria and cellular casts, decreases in proteinuria, and stabilization or improvement in renal function. A clinical trial intended to demonstrate induction of renal remission would specify a definition of renal remission that includes all relevant parameters. We recommend providing evidence supporting an association with improved clinical outcome (e.g., decreased likelihood of developing end-stage renal disease or need for dialysis) to defend the selected definition of renal remission. Because of concerns that patients with an inactive urinary sediment may nonetheless progress to renal failure, we recommend that studies using renal remission as an outcome measure include follow-up renal biopsies in at least a subset of patients.

   Patients with renal remission may be expected to experience a clinical benefit to the extent that they are: (a) spared treatment with potentially toxic agents; and/or (b) spared from ultimate progression to end-stage renal disease. We encourage sponsors proposing to use attainment of renal remission to demonstrate efficacy of a product for lupus nephritis to discuss their clinical development plans with the responsible reviewing division at the Agency. Proposals for clinical trials using renal remission as an endpoint should: (a) provide a clear definition for renal remission, and data supporting the choice of that definition; (b) provide evidence that attaining a renal remission would be expected to translate into a clinical benefit to the patient; and (c) assess the durability of the renal remissions.

5) *Resolution of nephrotic syndrome.* Patients with lupus nephritis may have high grade proteinuria with nephrotic syndrome. A clinical trial intended to demonstrate resolution of nephrotic syndrome would enroll patients with high grade proteinuria (e.g., ≥4 gm/d) and assess the proportion of patients who attain a prespecified, substantial reduction in proteinuria (e.g., to less than 500 mg per 24 hours). The trial should also collect data on the associated features of nephrotic syndrome (i.e., hypoalbuminemia, generalized

edema, and hyperlipidemia) to assess whether changes in these parameters mirror improvements in proteinuria. We encourage sponsors proposing to use resolution of nephrotic syndrome to demonstrate efficacy of a product for lupus nephritis to discuss their clinical development plans with the responsible review division at the Agency.

C. Complete Clinical Response/Remission

A complete clinical response/remission claim would be approved for products that demonstrate the ability to induce a clinical response, characterized by the complete absence of disease activity at all sites for at least 6 consecutive months. This response is termed *complete clinical response* if the subjects continue to receive lupus-directed therapies. Remission occurs if subjects were receiving no ongoing therapy for their SLE. A trial in support of the claim of *complete clinical response* should be at least 12 months in duration and demonstrate an increase in the proportion of subjects in whom a disease activity measure achieves zero.

D. Reduction in Flares

Reductions in the rate of flares of SLE or time to flare are considered to be clinically important outcomes. An increase in the frequency and severity of flares of lupus nephritis is correlated with worse outcomes. Thus, a reduction in the rate of flares of organ-specific disease (e.g., lupus nephritis) is also considered clinically important. If time-to-flare is evaluated as the efficacy endpoint, the study should be of sufficient duration to evaluate whether the flares are suppressed or only delayed in occurrence. Thus, a comparison of flare rate or incidence of flare-free at an appropriate time point will be a critical secondary endpoint. An established measure of flare may be considered in clinical trials studying flare as a primary outcome to demonstrate a decreased frequency of, or decreased severity of, flares. We recommend providing evidence that the chosen definition of flare accurately measures clinical flares. Proposals for clinical trials using renal flare as an endpoint should: (1) provide a clear and accepted definition for renal flare, and data supporting the choice of that definition; (2) provide evidence that reducing renal flare incidence by that definition of renal flare would be expected to translate into a clinical benefit to the patient; and (3) assess the durability of the renal benefit. A success in a clinical trial could be defined as an increase in the time-to-flare or as a decrease in the number or severity of flares over the course of a 1-year trial.

V. TRIAL DESIGN AND ANALYSIS

Careful consideration should be given to choosing endpoints that will accurately assess the clinical benefits of the product when designing a trial for SLE. The clinical trial can focus on one aspect of disease (e.g., lupus nephritis) over other important aspects. However, it is important to collect information about other aspects of disease to ensure an adequate assessment of the overall risk-benefit ratio. Clinical trials in SLE generally are expected to collect information about disease activity at all sites, irreversible damage due to SLE and its treatment, and valid HRQL measures. Serologic studies may also provide important information about the mechanism of action of the product under investigation.

A. Phase 2 Trials

Phase 2 trials are used to better define dose and exposure-related activity and toxicity of products under development. We recommend evaluating the safety of concurrent use of a new product with commonly used concomitant therapies, although at this stage studies will not be powered to adequately assess safety endpoints. Outcome measures under consideration for trials of SLE may not have been tested in large-scale randomized trials. Some outcome measures may prove less sensitive than expected. Unexpected confounding variables may complicate the interpretation of trials using these endpoints. Consequently, experience with these outcome measures in phase 2 trials can enable careful consideration to aid selecting valid, interpretable clinical outcome measures for the phase 3 trials.

B. Efficacy Trials

For the following discussion of efficacy trials in SLE, it is assumed that trials will be parallel arm, randomized controlled studies with a placebo or active control. Whereas in some trials the study drug will be evaluated as monotherapy, in many cases the study drug will be added to the standard therapy the patient was previously receiving (add-on trial). One of the advantages to an add-on trial of this type is that it allows the evaluation of pharmacokinetic and pharmacodynamic interactions with commonly used products in SLE. Alternative trial designs such as randomized withdrawal or replacement trials may also be considered. Investigators should discuss these alternative designs with the appropriate reviewing division before embarking on these studies.

1. Disease Activity Trials

For a clinical trial studying a reduction in disease activity, we recommend that the patient population to be enrolled reflect the patients who would reasonably be considered for this treatment should it be shown effective. It is important that the studied population be one that can be generalized to an appropriate population for recommended use, and not made artificially narrow. If existing data (e.g., from phase 2 studies) suggest that only a specific limited population is plausibly expected to benefit from the therapy, then the inclusion and exclusion criteria can limit enrollment to patients with a restricted range of disease activity. If the effects of treatment are expected to differ substantially in patients with severely active disease as compared to moderately or mildly active disease, then it may be desirable to stratify the randomization. Furthermore, in DAI trials, investigators may wish to stratify by organ to ensure balance between the two groups for at least one major organ system involved. In general, the indication statement in the package insert ultimately will reflect the patient population studied.

Clinical trials should be of sufficient length to assess the durability of benefits of therapy given the chronic nature of SLE and its waxing and waning course. Trials of 1-year duration are usually necessary (but see Section V.D.5., Trial Duration). One approach is to measure the effect on disease activity by comparing between groups the change in scores on a disease activity index between the outset and the end of the trial. Another approach is to use an AUC analysis based on disease activity assessments at regular intervals throughout the trial. An AUC analysis may more comprehensively measure disease activity during the study than at a single time point. However, AUC differences need to be interpreted carefully. Trials that collect outcome data at

multiple times during a trial can show the time course of treatment effects as well as intercurrent disease activity and thus better define the importance of the effect. Several confounding factors could complicate the interpretation of a trial that only examines baseline and study-end scores. First, many SLE patients have frequent low scores on disease activity indices, but experience intermittent flares of disease. A study examining only study-end scores may be insensitive to the benefit of a new product which decreases the frequency and severity of disease flares but has only a small effect on background disease activity. Another confounding factor is the likelihood that subjects who flare during the trial will be treated with additional medications (e.g., corticosteroids), potentially reducing their disease activity scores for reasons unrelated to the study drug (see also Section V.D.1., Concomitant Medications).

In a clinical trial intended to show an improvement in a DAI, it is important to ensure that the outcome measure accurately assesses disease activity in the treated patients. Some disease activity indices give points for a new disease manifestation and no points for a stable manifestation. Thus, a disease manifestation that is present at screening that is stable during the study could contribute points to the baseline score but no points to subsequent scores leading to an artifactual reduction in the overall disease activity score. We recommend the protocol include definitions of disease manifestations, and levels of disease severity be clearly specified. The interpretation of score changes may be confounded if organ system dysfunction due to a disease or condition other than SLE is present, or organ dysfunction due to the treatment occurs. It is important that the study protocol specify procedures to ensure that the scoring of the DAI specifically reflects SLE-related organ dysfunction. Clearly, there are situations when changes in scores may not accurately reflect changes in disease activity. These limitations do not preclude the use of these disease activity indices in clinical trials, but the investigator should be aware they exist. In addition, careful training of investigators is essential to ensure uniform scoring. If there is a lack of reproducibility of these measures from clinician to clinician, it may seriously impair the interpretability of the trial results.

We recommend analyzing the results of clinical trials to verify that an improvement in a disease activity score represents a clinical benefit to the patient and to assess the generalizability of the results. It is important that patient outcomes be analyzed to determine that the improvement in disease activity is not accompanied by worsening in other disease manifestations. Overall, assessment of irreversible organ damage defined as histologic or functional changes and/or measures of HRQL should not significantly worsen. To explore the generalizability of the benefits seen, we recommend subset analyses be carried out regarding the extent of benefit for disease affecting specific organ systems.

Another method to measure a decrease in disease activity is to assess the incidence of disease flares during the course of a clinical trial. This type of trial might use measures of mild/moderate and severe SLE flares as the primary outcome measure. As not all SLE patients experience flares in a given time frame, the size and duration of the trial should be adequate to capture a sufficient number of flares in the treatment and control groups to demonstrate a decrease in the treatment arm. Collection of complete information on concomitant medications is essential to ensure that a difference in the number of SLE flares is attributable to the study drug. We recommend careful consideration be given to determining the appropriate regimen for the control arm of a trial in SLE. No subject should be denied recognized effective treatment for aspects of

the disease which may lead to irreversible harm. A design consistent with this principle randomizes subjects to the addition of placebo or study drug to a generally acceptable standard of care regimen. This seeks to demonstrate that disease activity is decreased in the treated subjects. A study could also randomize subjects to the receipt of a known active agent or the study drug, then assess if there is a larger decrease in disease activity in subjects receiving the new product. It may be appropriate to include early escape provisions for subjects who worsen during the study to ensure that no subject is denied potentially effective therapy.

2. Lupus Nephritis Trials

Measurement of renal disease in SLE in clinical trials requires knowledge of the histologic description delineating the extent of inflammation or scarring, because the outcome and clinical features vary markedly among the various WHO categories of lupus nephritis. A variety of endpoints can be used to demonstrate efficacy in lupus nephritis, including progression to end-stage renal disease, progression to a specified level of loss of renal function as assessed by serum creatinine or creatinine clearance, induction of renal remission, reduction in renal flares, and resolution of nephrotic syndrome. A discussion of the use of these endpoints in clinical trials is provided in Sections III.C. and IV.B. and D.

3. Other Organ-Specific Claims

Responder measures for each organ system studied can be proposed and based on organ-specific measures from a DAI. If an organ-specific outcome is studied, we recommend a comprehensive DAI be included as a secondary outcome. A responder measure has the advantage of addressing the particular disease manifestations of most concern for an individual patient. This approach recruits a more homogeneous population of patients compared to the DAI approach, although it is recognized that patients will often have more than one organ system involved. Powering such a study may be problematic if study enrollment is restricted to patients with one specific organ system involved. Patient populations with disease affecting more than one organ can be studied using an organ-specific approach if the organ system or systems that have been most problematic for each enrolled subject are identified. Trials can study a single organ or they might study disease in more than one organ, with stratification by each patient's primary organ of involvement, allowing evaluation of effects on several specific organs within a single trial. Stratification by extent of organ damage at baseline may be advantageous to ensure balance of pre-existing organ damage between treatment groups. We recommend that clinically important outcomes be defined for each organ system, and composite endpoints can be considered. In disease activity trials, we recommend measuring multiple time points, which can improve efficiency of the trial.

A successful trial may demonstrate a statistically significant number of clinical remissions in the treated group versus the control group. Trends for improvement in each organ system can then be examined. However, the interpretation of a clinical trial using the specified organ approach could be problematic if worsening in other manifestations of lupus counterbalanced improvement in the organ system measured. If changes in treatment regimens are made, such as an increase in immunosuppressive agents, the results in the designated organ would be confounded.

C. Studies to Show Superior Safety

Studies to demonstrate the improved safety profile of a new drug compared to standard therapy may also be considered. We recommend these trials also be of adequate duration to establish efficacy. If comparable efficacy is expected, rather than superior efficacy, then a noninferiority design to evaluate efficacy will be necessary. Rigorous noninferiority demonstrations are necessary, but can be difficult to achieve. It is recommended that sponsors proposing such studies identify the known effect size for the comparator and define a noninferiority margin that preserves a sufficient percentage of the effect size to demonstrate efficacy with the new product. These choices must be based on careful and comprehensive review of the data available regarding the comparator agent. It is also important for these studies to be powered to demonstrate that the new product is noninferior and to adequately assess the claim of an improved safety profile. It is appropriate for steroid sparing agents to demonstrate not only that reduction in steroid use is statistically significant, but also that these reductions translate into an improved safety profile. Ensuring that a trial has sufficient power to demonstrate improved safety may be problematic in lupus, although studying a collection of important adverse events may help in this regard. Other trial designs may be considered but it is recommended that these be discussed with the appropriate reviewing division before initiation.

D. Other Trial Design Issues

1. Concomitant Medications

We recommend careful consideration of the use of concomitant medications during trials. This includes defining allowable medications at baseline and allowable changes in medications during the trial. It is important that investigators consider restricting baseline glucocorticoid use (stable dose or limit the range of doses) to reduce the variability of dosing that may introduce bias and make interpretation of results more difficult because of significant variation and imbalances of initial doses. If glucocorticoid dose changes are allowed during the trial, it is important that these changes be carefully discussed in the protocol before the trial begins. We also recommend considering the use of rescue medication and whether patients requiring rescue medication be withdrawn from continued administration of randomized study agent. It is important to recognize that subtle changes in concomitant medications, whether steroids, immunosuppressive agents, or other therapies, can influence outcomes. It is important for the protocol to provide consideration for standardization to the use of concomitant medications including ACE inhibitors and antihypertensive agents, levels of blood pressure, and control of diabetes (especially for studies of lupus nephritis).

2. Issues of Blinding

Blinding is intended to minimize the potential biases resulting in differences in management of patients or assessment of patient status. Therefore, it is important that every effort be made to ensure that trials are adequately blinded. This can require, among other things, identification of third parties to assess efficacy, to administer drugs, or to make patient management decisions.

637 *3. Standard of Care Issues*
638
639 No patient enrolling in a clinical trial should be denied standard therapy if that may lead to
640 irreversible harm. To avoid denying patients standard of care, clinical trials of new therapies can
641 use add-on study designs, or head-to-head comparisons with an alternative standard of care.
642 Corticosteroids with or without cyclophosphamide plus placebo compared to corticosteroids with
643 or without cyclophosphamide plus new drug is an example of an add-on design that assesses
644 efficacy of a new product as compared to placebo in the context of background corticosteroids or
645 corticosteroids plus cyclophosphamide.
646
647 To the extent that cyclophosphamide may be effective, demonstration of an effect of a new drug
648 may be difficult in trials in which cyclophosphamide is considered part of the standard of care
649 regimen, especially if the mechanisms of action of cyclophosphamide and the new therapy are
650 similar. It may be difficult to identify toxicity of the new drug in the context of the use of
651 multiple immunosuppressive agents. We recommend that sponsors consider these issues when
652 designing trials.
653
654 *4. Extension Trials*
655
656 Extension trials are used to demonstrate maintenance of efficacy observed in a short-term
657 evaluation, and long-term safety. We recommend that sponsors consider whether comparators
658 are warranted in these studies, and whether these extension studies be blinded or open label.
659 Although it may be difficult to perform a blinded extension study, advantages to this include
660 obtaining more robust efficacy and safety data. The more robust nature of the data can be
661 important to weighing the strength of the evidence in making risk-benefit comparisons, and
662 achieving claims in approved labeling.
663
664 *5. Trial Duration*
665
666 In general trials should be 12 months in duration although trials of shorter periods can be
667 considered, depending on the organs and outcomes studied. Short-term trials may not provide
668 adequate demonstration of efficacy, safety, and durability of response. However, it may be
669 difficult to perform long-term studies secondary to flares, changing medications, dropouts, and
670 changes in medical practice.
671
672
673 **VI. SURROGATE MARKERS AS ENDPOINTS**
674
675 Surrogate or early markers of disease activity can be considered for assessment of efficacy in
676 lupus trials. Such markers can be particularly useful in phase 2 studies, prior to definitive
677 demonstrations of efficacy. If surrogate endpoints are being considered for the demonstration of
678 efficacy to support a marketing application, we recommend they be thoroughly discussed with
679 the FDA reviewing division and be validated for the treatment under study. Approval may be
680 based on a validated surrogate endpoint. If the surrogate is not validated, but appears to be
681 reasonably likely to predict a clinical benefit, accelerated approval may be considered under 21

CFR 314, subpart H or 21 CFR 601, subpart E. In this case, approval would be contingent upon a phase 4 study to verify the clinical benefit.

Supporting the proposition that the surrogate is reasonably likely to predict clinical benefit is essential to this approach. An effect on the surrogate should be demonstrated in adequate and well-controlled clinical trials. Trends toward clinical improvement observed in the trials that establish an effect on the surrogate marker can serve to strengthen an assessment of the surrogate as being reasonably likely to predict clinical benefit. The totality of the available data will be examined during the review process in considering a product for accelerated approval. The ability of the surrogate endpoint to predict clinical outcomes will be weighed against the risks associated with treatment.

Potential surrogate markers can be laboratory evaluations involving physiological indicators or pathological changes identified in the organ under study. For example, a sustained doubling of serum creatinine is a valid surrogate marker for the clinically important outcomes of ESRD, and the need for dialysis or renal transplantation. Changes in creatinine clearance or iothalamate clearance can also be considered as potential surrogates for ESRD. Significant changes as assessed by repeat renal biopsies also have potential to serve as a surrogate endpoint. A significant improvement in hematuria and proteinuria in conjunction with a substantial change in the level of anti-double-stranded DNA antibodies can be proposed for consideration as the basis for approval. Other composite surrogates can also be considered. Other markers might include assessment of B- and T-cell subsets, autoantibody subsets, immune complexes which are specifically defined, presence or absence of procoagulants, complement or its products. It is possible that *proof of concept* studies can be useful to support subsequent designs leading to consideration of approval. For example, sponsors can consider measuring the effects of a study drug against the effect of true placebo on T- and/or B-cell profiles in short-term trials to determine a measure of potential efficacy, possible dose, and treatment duration for subsequent study in pivotal trials for approval. However, to be suitable as a basis for accelerated approval, it would be appropriate to have strong evidence that the proposed surrogate is *reasonably likely to predict clinical benefit*. We recommend sponsors be cautious about selecting a surrogate endpoint intended to support accelerated approval until there is confidence regarding its predictive value.

VII. RISK-BENEFIT ASSESSMENT

Approval of a therapy for SLE is predicated on evidence from adequate and well-controlled studies demonstrating efficacy and safety that support a conclusion of an acceptable risk-benefit. Assessment of risks and benefits requires an appraisal of the impact of the product on all aspects of the disease process, including disease activity, irreversible damage due to SLE and its treatment, and quality of life (Strand 1999). It is important that the size of the safety database at approval be consistent with the recommendations made by the International Conference on Harmonisation (ICH guideline E1A).[3] Particular attention should be paid to the assessment of known toxicities, or to pharmacologic effects that might be suspected to imply delayed toxicities.

[3] ICH guideline for industry *E1A The Extent of Population Exposure to Assess Clinical Safety: For Drugs Intended for Long-Term Treatment of Non-Life-Threatening Conditions*

It is important to consider these toxicities in formulating the clinical development program and this may influence the size of the necessary safety database. The recommended size of the safety database may be lower for orphan indications, as it may be impossible or impractical to study a large number of subjects. Although SLE is not an orphan indication, there may be subsets of patients with specific manifestations of SLE who represent an orphan population indication. Sponsors may wish to discuss these issues with the appropriate FDA staff early in the development of a new treatment. Finally, if there is concern about rare but serious adverse events (e.g., from the mechanism of action or experience with similar agents), a phase 4 commitment may be needed to gather additional safety information.

VIII. LUPUS AND PHARMACOKINETICS

A. General

For many products there have been few pharmacokinetic studies done in a prospective manner in the lupus population. The bulk of the pharmacokinetic experience in these subjects has been anecdotal in nature. However, pharmacokinetic data may serve an important role in designing the clinical development program. For example, determining the dosing interval of a drug in individuals with lupus may be a challenge because of the multisystem nature of the disease. It is important that patient enrollment in pharmacokinetic studies reflect the population for which the drug is intended. As women represent the primary population afflicted with lupus, we recommend that enrollment in pharmacokinetic studies incorporate a preponderance of women. Due to the multisymptom and body system nature of lupus, it is important that subjects enrolled in pharmacokinetic trials for lupus have organ system involvement to assess the need for organ-specific recommendations.

B. Special Studies

A characteristic feature of lupus is the associated change in the kidney, both structurally and functionally. These kidney changes make it difficult to determine whether the standard renal transplant model is adequate for the assessment of declining renal function in the lupus patient. It is recommended that separate pharmacokinetic trials be considered in lupus patients with varying degrees of proteinuria to assess the impact on drug disposition and binding (e.g., those with proteinuria greater than 4 grams/24 hours, greater than 1 gram/24 hours, or greater than 500 mg/24 hours).

C. Drug Interactions

We recommend conducting drug interaction trials with those agents commonly used in the treatment of lupus. It is important to assess the potential for interactions with hormonal contraceptives. These assessments can include either in vitro or in vivo methodologies or a combination. The reader is directed to the published FDA guidances on in vivo and in vitro drug interaction studies (see References).

REFERENCES

Boumpas, DT and JE Balow, 1998, Outcome Criteria for Lupus Nephritis Trials: A Critical Overview, Lupus, 7:622-629.

Food and Drug Administration, 1997, Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro, April 1997.

_____, 1999, In Vivo Drug Metabolism/Drug Interaction Studies — Study Design, Data Analysis, and Recommendations for Dosing and Labeling, November 1999.

Levey, AS, SP Lan, HL Corwin, BS Kasinath, et al., 1992, Progression and Remission of Renal Disease in the Lupus Nephritis Collaborative Study: Results of Treatment with Prednisone and Short-Term Oral Cyclophosphamide, Ann. Int. Med., 116:114-123.

Petri, M, M Genovese, E Engle, and M Hochberg, 1991, Definition, Incidence, and Clinical Description of Flare in SLE. A Prospective Cohort Study, Arth. Rheum., 34:937-44.

Strand, V, D Gladman, D Isenberg, M Petri, J Smolen, and P Tugwell, 1999, Outcome Measures to Be Used in Clinical Trials in Systemic Lupus Erythematosus, J Rheumatol, Feb;26(2):490-7.

Stoll, T, B Seifert, and DA Isenberg, 1996, SLICC/ACR Damage Index Is Valid, and Renal and Pulmonary Organ Scores Are Predictors of Severe Outcome in Patients with SLE, Br. J. Rheumatol, 35:248-54.

Urowitz, MB and DD Gladman, 1999, Evolving Spectrum of Mortality and Morbidity in SLE, Lupus, 8(4):253-5.

APPENDIX: GLOSSARY OF ACRONYMS

AIMS	Arthritis Impact Measurement Scale
AUC	Area Under the Curve
BILAG	British Isles Lupus Assessment Group
CDER	Center for Drug Evaluation and Research
DAI	Disease Activity Index
ECLAM	European Consensus Lupus Activity Measure
ESRD	End-Stage Renal Disease
FDA	Food and Drug Administration
GFR	Glomerular Filtration Rate
HAQ	Health Assessment Questionnaire
HRQL	Health-Related Quality of Life
ICH	International Conference on Harmonisation
KFSS	Krupp Fatigue Severity Scale
MHAQ	Modified Health Assessment Questionnaire
SELENA	Safety of Estrogen in Lupus Erythematosus National Assessment Trial
SLAM	Systemic Lupus Erythematosus Activity Measure
SLE	Systemic Lupus Erythematosus
SLEDAI	Systemic Lupus Erythematosus Disease Activity Index
SLICC/ACR	Systemic Lupus Erythematosus International Collaborating Clinics/ American College of Rheumatology
WHO	World Health Organization