CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: NDA 20905

MEDICAL REVIEW(S)

MEDICAL OFFICER REVIEW
LEFLUNOMIDE
NDA 20,905

MESSAGE TO READERS

This review is a discussion of the major aspects of the analysis of efficacy and safety for Leflunomide. In the EFFICACY REVIEW the dosing decision process is briefly reviewed, as all three major clinical trials using the same 20mg/day regimen. The protocols themselves are then summarized, followed by an important section on methodolgy regarding both design and analysis of RA trials investigating both clinical and radiographic endpoints. The studies are homogeneous regarding endpoints and primary analyses, but the details of the analyses are critical. Data on other endpoints are presented, along with exploratory work on the association of clinical and radiographic outcomes, and data from the small pharmacokinetic / clinical study of the combination use of leflunomide and methotrexate in RA. The SAFETY REVIEW uses a conventional organizational format.

All clinically relevant data are translated and distilled into labelling language. The clinical implications of the considerable pharmacologic and pharmacokinetic data in this submission are not elaborated here, but their implications for labelling are evident in the respective sections of the draft.

EFFICACY REVIEW

DOSING DECISIONS

The decision of 20mg/d of LEF was made on the basis of phase 2 dose ranging, clinical, population pharmacokinetic study (YU203). This 6 month, double blind, placebo-controlled safety and efficacy study showed statistically significant differences between the 10mg, and 25mg/d doses compared with placebo in all four traditional RA efficacy parameters (tender and swollen joint counts, investigator and patient globals) as shown below.

Table 30 Baseline Means and Mean Changes in Joint Counts and Global Assessments in YU203

Variable		Mean (SD) – Intention-to-Treat Analysis							
		Placebo (N = 102)		5 mg LEF (N – 95)		10 mg LEF (N = 100)		25 mg LEF (N = 101)	
Tender joint count	Baseline	19.9	(6.59)	19.1	(6.12)	20.1	(6.12)	19.6	(6.57)
	Change	-4.4	(6.66)	-5.2	(6.78)	-7.9	(8.33)	-9.5	(8.11)
	p-value	–		0.3444		0.0011		<0.0001	
Swollen joint count	Baseline	15.9	(6.30)	15.4	(5.52)	16.1	(5.60)	15.6	(5.81)
	Change	-3.6	(6.04)	-4.4	(6.44)	-6.1	(5.99)	-7.3	(5.91)
	p-value	–		0.2220		0.0034		<0.0001	
Inv. global assessment	Baseline	6.1	(1.64)	5.8	(1.64)	6.0	(1.50)	5.8	(1.61)
	Change	-1.4	(2.53)	-1.8	(2.42)	-2.7	(2.54)	-2.7	(2.61)
	p-value	–		0.0816		<0.0001		<0.0001	
Pat. global assessment	Baseline	6.4	(1.98)	6.0	(1.83)	6.1	(1.86)	5.8	(1.77)
	Change	-1.3	(3.04)	-1.5	(2.90)	-2.7	(3.04)	-2.6	(2.61)
	p-value	–		0.0794		<0.0001		<0.0001	

REF: Appendix Tables 25-28

The population PK component indicated a correlation of a positive clinical effect of leflunomide and a total plasma concentration of its active metabolite (A77-1726) of at least a steady state concentration of 13mg/L. A population kinetics analysis then showed that a daily dose of 20mg/d was necessary to achieve this in 99% of patients. Multiple dose PK analyses showed that a loading dose of 100mg/d for 3 days results in the rapid achievement of this plasma level.

Finally a dose response analysis of the YU203 data was done to predict the clinical

response by joint counts to a 20mg/d dose. These results are shown below, indications clear separation of the 95% confidence intervals of the 20mg/d and 25mg/d doses (and, to a lesser degree, the 10mg/d dose) from that of placebo.

Table 31 Results of Efficacy Analysis: Tender Joint Count		
Dose	Predicted	945% CI
0	-9.8	-11.7 to -7.8
5	-11.2	-12.7 to -9.7
10	-12.6	-13.9 to -11.2
20*	-15.4	-17.0 to -13.7
25	-16.8	-19.3 to -14.3

REF: Executive Summary, Phase II Studies
*Calculated, not observed
Note: Regression model: Change = -9.77 + (-0.280) (Dose); p≤0.0001

Table 32 Results of Efficacy Analysis: Swollen Joint Count		
Dose	Predicted	95% CI
0	-7.0	-8.4 to -5.6
5	-8.0	-9.1 to -6.9
10	-9.1	-10.0 to -8.1
20*	-11.1	-12.3 to -10.0
25	-12.2	-14.0 to -10.4

REF: Executive Summary, Phase II Studies
*Calculated, not observed
Note: Regression model: Change = -6.99 + (-0.207) (Dose); p≤0.0001

APPEARS THIS WAY ON ORIGINAL

EFFICACY OVERVIEW

The leflunomide NDA was submitted on March 10, 1998, and includes three pivotal trials of clinical and radiographic endpoints: US301 - a one year comparison of leflunomide (LEF), methotrexate (MTX) and placebo (PLC), MN301 - a 6 month comparison of LEF, sulfasalazine (SSZ), and PLC, with an optional, blinded 6 month extension of LEF and SSZ arms, and MN302 - a one year comparison of LEF and MTX. All trials are being continued blindly for two years. The trials tested two hypotheses, although the hypotheses cannot be considered entirely equivalent. Trial design was, in the first instance, clinically driven, because RA treatment decisions are currently much more based on clinical than radiographic considerations. Protocol design, implementation, and analysis for the three pivotal trials in this NDA has been an ongoing, interactive process over the past several years, aiming to yield analyses least vulnerable to ambiguous or controversial interpretation. The full protocols and amendments will be available; a summary description of the protocols and analyses are given below.

The data successfully demonstrate "reduction in signs and symptoms" (one of a number of claims offered in the RA Guidance Document (3/1998), attached). The radiographic data also demonstrate "prevention of structural damage" by protocol defined analyses, but assuring that missing data do not undermine this demonstration is more challenging than with the "signs and symptoms" claim. The NDA also addresss a "prevention of disability" claim, which earlier had entailed one year evidence, but now consists of two years data. This claim focuses on physical function, with the additional proviso that health related quality-of-life (HR-QOL) "not worsen". Nonetheless, these data are presented because they have a certain intrinsic persuasiveness. They will constitute part of a later submission of two year data. Pharmacoeconomic data were also collected, which have not been submitted.

PROTOCOL SUMMARIES

US301

This was a one year US, multicenter, randomized, double-blind comparison of the clinical and radiologic efficacy of leflunomide (LEF) 20mg/d, methotrexate (MTX) 7.5-15mg/wk, and placebo (PLC) in 482 RA patients with disease at least 6month duration, ACR "active", and MTX naive, assigned with a 3:3:2 ratio, and stratified by time from last disease-modifying-antirheumatic-drug (DMARD) of greater than 8weeks or not. It was thus both a difference design - LEF vs PLC, and an equivalence design - LEF vs MTX, using the agreedupon test that the lower limit of the 95% confidence interval of the difference of the two active arms exceed minus

10%. All confidence intervals noted in this review are 95% confidence intervals. If the 95% confidence interval lies fully to the right of zero, it can also be concluded that LEF is statistically significantly superior to the active control. This protocol also specified that patients not demonstrating an ACR20 response on or after 4 months or with adverse reactions requiring withdrawal could (blindly) change treatment (after 4 or 8 week washouts as, if indicated): MTX or PLC changed to LEF, LEF changed to MTX. All patients received folate at 1mg BID.

The primary clinical analysis was a comparison of proportions of ACR20 responding patients at 12 months; all other patients were classified as failures. The mean AUC duration (in weeks) under the "ACR20 response curve" was the secondary analysis. A positive result here, as in all efficacy analyses in this NDA, may be confounded, or even negated, by a very skewed (differential) dropout pattern, so dropout analyses are necessary (see Methodology, below) to ensure this did not occur. The primary radiographic analysis was a comparison of mean change in x-ray Sharp scores (see glossary for definition of Sharp score) from baseline to end of trial (regardless of intervening drug changes). The secondary radiographic analysis was Sharp score changes from baseline to dropout (ie. termination of initial assigned drug) by "intention-to-treat, last observation carried forward" (ITT/LOCF) for imputing missing data.

MN301

This was a 6month, multinational, multicenter, double-blind comparison of LEF 20mg/d, sulfasalazine (SSZ) 2gm/d, and PLC in 358 RA patients assigned with a 3:3:2 ratio and stratified on disease duration of less than 2years or not. An optional 6month extension, continuing the blind, was avaliable for 6month completers. In the NDA this extension was called MN303, and it was elected by 80 of 96 LEF, 76 of 83 SSZ; and 41 of 51 PLC patients (who, per protocol were to be switched to SSZ). Thus, this protocol is an acceptable 12 month equivalence design (LEF vs SSZ), provided the 6 month comparison of the active control (SSZ) validates the assay by demonstrating superiority to PLC, and there cannot be demonstrated a differential recruitment effect with patients enrolled into the second six month component. MN301 is thus both a difference and an equivalence design. The same allowable maximal small difference of 10% for an equivalence success was used here as in US301. The MN301 protocol required dropout for certain toxicities and for inefficacy defined as: 3 of the following - fall in tender joint count of 2 or less, fall in swollen joint count of 2 or less, no improvement in patient global, and no improvement in investigator's global.

The clinical analyses were identical to US301: primary being a comparison of proportions of ACR20 responders at trial completion, and secondary the mean AUC under the ACR20 response curve. X-rays were obtained at baseline, 6 months, and

12 months here; there was no xray at dropout point. Thus, the radiographic analysis was simply a comparison of the mean change in the x-ray Sharp scores from baseline to 6 month, and, regarding the second six month component, a baseline to 12 month, and a 6 month to 12 month comparison.

MN302

This was a 12 mo multinational, multicenter, double-blind comparison of LEF 20mg/d and MTX 7.5-15mg/wk in 999 RA patients with disease less than 10 years duration, assigned using a 1:1 ratio of 999 patients. No protocol specified inefficacy dropout criteria were included in the protocol. This was an equivalence design, using the same "equivalence test" as in the above trials. The clinical analyses were identical to trials US301 and MN301 above; and the radiographic analysis as with MN301 because only baseline and 12 month x-rays were obtained.

METHODOLOGY

EFFICACY ASSESSMENTS

The RA Guidance Document suggests two preferred clinical measures, the ACR20 response and the "traditional 3 of 4 endpoints": tender joint count (TJC), swollen joint count (SJC), patient global (PG), and investigator global (IG). The ACR20 response is used in these trials; the "traditional 3 of 4" data and other data are also presented.

The RA Guidance Document describes an enhanced claim of "prevention of disability" for a product which is durable for at least two years (trial duration minimum) and impacts patients in important functional ways. The claim requires the co-demonstration of no worsening in either a disease specific health-related quality-of-life (HR-QOL) measure or a generic HR-QOL measure that has been shown sensitive in RA (see RA Guid. Doc.). All three trials assessed physical function with the modified HAQ (MHAQ). Trial US301 also used the Problem Elicitation Technique (PET), a preference disability questionaire, and this trial also assessed HR-QOL with the SF-36. The PET incorporates patient individual preferences by inquiring what aspect of their arthritis they would most like to improve, and so this measure is arguably the optimal method to assess disability in arthritis.

For radiographic assessment the Guidance Document suggests the Sharp measure (or the Larsen, although recent work (OMERACT, '98 - to be published) seriously questions whether the Larsen should be considered as good as the Sharp) for radiographic assessment. The Sharp score is the cumulative score obtained by

assaying on a 0-3 scale joint space narrowing (JSN) and erosions in all hand/wrist and forefoot joints. The maximal Sharp score is 422. Studies have show that the Sharp instrument demonstrates acceptable intra-observer variability over time for trained readers (and this NDA used Dr. Sharp himself to read the films). Since the Sharp method is proving more discriminating (see OMERACT, '98), only the Sharp scores will be used here. Information on differential utility of erosions versus JSN, or hand/wrist versus forefoot films, will be useful to rheumatologists, but do not add further weight of evidence.

TRIAL DESIGN WITH CLINICAL AND RADIOGRAPHIC HYPOTHESES

Given the limited knowledge of the clinical utility of xray changes in RA patients, trial design needs, in the first instance, to be "clinically driven". Thus, a priori, designs for clinical hypotheses are more robust than those for radiographic ones. The biggest trial design problem with xray trials is that of "sparse data" compared with the abundance of clinical measurements possible. The biggest trial conduct problem is the number of patients with missing or unreadable xrays. This latter reflects more non-compliance in the xray realm than in the clinical. Patients sometimes feel the xray itself involves some risk. These factors make xray trial design potentially more uncertain at this time.

The above imbalance, however, is NOT the reason that, at this point in time, affirmation of a radiographic hypotheses must be accompanied by a clinical demonstration. The reason for this is simply our ignorance as to whether there, in fact, exists a clinically meaningful association between worsening in signs and symptoms and a worsening of the xray by erosion and JSN assessment. Although in the extreme, clinical-radiographic correlation is evident -- a normal xray in an asymptomatic person, or an endstage xray in a debililated one, what correlation exist in general is unknown. For this reason, and because even a known correlation epidemiologically does not mean beneficially altering the radiograph in an interventional trial automatically translates into clinical benefit. Thus, the "improvement in structure" claim normally needs a concommitant clinical claim (see RA Guidance Document).

However, since xray damage in RA is largely irreversible, it has a cumulative dimension (after the time to onset of the treatment) which is not a feature of the traditional clinical measures. In some sense, the xray is an AUC of damage over time, so it is arguably a preferred endpoint for longer term assessments. It can reflect, in patients on sequential therapy, effects of all drugs, weighted by exposure adjusted for time to onset.

The question of what extent clinical changes correlate with radiographic ones has never been addressed in any formal and robust venue, and this database will allow this exploration, but the development plan could not directly address it. The trial designs used here had to be primarily "clinically driven". To have done otherwise would have not been ethically justifiable. Only if it were already known that xray changes strongly and reliably predict prognosis could one justify an "xray driven design", eg. using xrays to determine change of treatments.

Nonetheless, the data will shed light on the relationship between the clinical and radiologic response, and this information may even merit description in the label. Also, the data was analyzed using two *a priori* by-patient radiographic criteria: (1) ***patients with no newly-eroded joints***, ie. no joint being non-erosive at baseline then showing a clearly identificable erosion at endpoint, and (2) ***patients with Sharp score worsening of less than four*** (this being, from Dr. Sharp's past experience, the upper limit of variability in repeated readings of the same film).

METHODOLOGY RATIONALE

CLINICAL EFFICACY

The primary clinical analysis for all trials is the proportion of ACR20 responders at trial end (6 or 12 months). This definition was selected for analytic simplicity and clinical understandability; it is a variant of a survival design, benchmarked to a specific timepoint. Excepting very anomalous situations, the primary analysis should be replicated by the secondary analysis, the mean AUC duration under the ACR20 response curve. This addressed the importance placed on an "all timepoint analysis" (see RA Guidance Document), and is the mean cumulative time throughout the trial (which may be comprised of discontinuous segments) during which the patient showed an ACR20 response. Other data analysed were the "traditional 3 of 4", and the percent ACR50 and ACR70 responsers. Analysis of missing clinical data is approached in the usual ways (see below).

MISSING CLINICAL DATA

Since it is unwarrented to assume missing clinical information (due to-dropouts or noncompliance) is "completely-at-random" in these trials, it is necessary to analyze dropout behavior to refute the assertion that ***an inference is due to dropouts, not drug effect***. This is an issue in any arthritis trial, unless its inference holds desp;ite application of the "worst-case scenerio", wherein maximal scores are assigned and carried forward for placebo dropouts and minimal scores for test drug dropouts. Reflection shows that, except where there are very few dropouts or overwhelming treatment effects, the worst-case approach negates all treatment inferences.

Unfortunately, no analysis of dropout behavior can be derived from first principles, and modelling, despite its theoretical attractiveness, cannot, in the end, satisfy this concern, because data needed to verify the model's assumptions are not available because they are ***missing***. In this review a number of dropout analyses are done, and if enough (non-redundant) analyses fail to support the assertion that differential dropout behavior is responsible for the outcome, the assertion can be assumed refuted. To some degree this will always be a judgment call.

1. Comparison of proportions of dropouts by category -- inefficacy, toxicity, "other" (adjusted for baseline differences as indicated). Comparisons of time-to-dropout for each category by logrank.

These are a traditional and straightforward, although even here simple comparisons may be complicated by patients who have more than one reason for exiting a trial. Nonetheless, certain patterns are expected, eg. more inefficacy dropouts in placebo arms, or more toxicity dropouts in active arms. Anomalies often arise, however, such as arthritis trials showing drug patients tolerating small improvements less well than placebo patients because the aggravating effect of some low grade but ongoing toxicity (such as nausea) makes them more inclined to dropout. If this occurred in an active control but not in the test drug arm, a differential dropout pattern would result.

2. Scattergrams of efficacy measures at dropout time, and statistical tests of whether systematic differences across treatment arms exist.

This would uncover the phenonena noted above. However, even here, incorrect inferences might still result from erroneous assumptions about the therapy. An illustration is the clinically suspicion (probably true) that RA patients responding very positively to injectable gold therapy often herald that response with a significant, but transient, gold toxic reaction (eg. rash). In a standard RA trial these patients may dropout for toxicity and so prevent detectiing efficacy. Only close followup of the dropouts would reveal this anomaly.

3. Formal followup of dropouts.

This will reveal the presence of post-dropout drug effects. This goal was variably achieved in the three pivotal trials of this NDA. For US301, followup of patients failing initial therapy was protocol formalized, with (blinded) switching to another active agent.

RADIOGRAPHIC EFFICACY

The conceptual approach to efficacy analyses of radiographic data is, in principle, the same as for the analysis of clinical efficacy, but the "sparse data" aspect makes the application more difficult. The goal of any efficacy analysis is to capture all timepoint

data on all randomized patients. Equally fundamental, in any drug registration trial, is the goal of an inference regarding the drug – ie. "drug attribution". Analyses need to be done on patients on the assigned drug, and to the degree that this is achieved is the extent that the inference applies to the experimental drug. However, a problem immediately arises because these two goals, fully expressed, are mutually incompatable, unless there are zero dropouts or unless there is a persuasive argument that dropouts are "completely at random", neither of which probably ever occur in arthritis trials.

Given this, there are, a priori, two approaches to radiographic data analysis: (1) analyze "time on all drugs" and, separately, consider effects of alternative therapies, and (2) analyze "time on initial drug" and, separately, considering follow up of dropouts on alternative therapies. Analyzing by "time on all drugs" inevitably means confounding by other (at least potentially) active agents, as would occur if a patient dropped out a LEF arm and was switched, for instance, to MTX. However, this analysis does have the advantage of assuring no loss of patient followup, and it is the ***only*** analysis possible if no dropout point xray was allowed (as in trials MN301 and MN302). Thus, this was used as the prmary radiographic analysis. The three trials in this NDA called for baseline and end-of-trial xrays for all randomized patients, provided they received at least one month of therapy. The end-of-trial xray captures intervening disease activity for a "time on all drug" analysis, even if the patient discontinued his originally assigned therapy. The US301 trial, additionally, called for an xray at dropout point when applicable; European investigators felt designs using a third xray was excessive radiation exposure and hence ***unjustified***. The US301 trial permitted a secondary radiographic analysis by "time on initial drug", using the dropout xray score carried forward. The is a standard intent-to-treat/last observation carried forward analysis.

MISSING RADIOGRAPHIC DATA

In principle, missing radiographic data present the same question of their effect on the inference as do missing clinical data, but the consequence of one missing (or unreadable) xray is much greater than one missing clinical datapoint because the entire contribution of the patient to the xray analysis is lost. There are only two films done over the entirety of the trial (or, for dropouts in US301, three), unlike the monthly clinical datapoints. On the other hand, a differential dropout pattern, a priori, would seem less likely with xrays than with clinical data because neither patient nor investigator usually would know the xray score at the dropout decision point. In fact, in trials MN301 and MN302, this score was unknowable.

The traditional approach to analyzing missing xray data would be to look at the missing data patients across arms to see if they markedly differed clinically. This was done and is included in the review, but it has limited persuasiveness because of the very small numbers of patients in the various categories of dropouts.

Another approach, a priori, would be to do sensitivity analyses to see how deviant the missing data cohorts results could be and still have the overall inference hold (to $p<0.05$). This "maximal deviance tolerated by the data" is a function of both (1) the mean difference in the drug cohort with missing xrays and the placebo cohort with missing xrays, and (2) the variability difference between the same two cohorts. For ease of understanding by clinicians, I elected to determine what maximal mean difference could be so tolerated, fixing the variability. Accordingly, this was done for the two placebo-controlled studies (US301 and MN301). Thuis model employing the assumption, which does not seem overly burdensome, that the variability of missing data is similar to the variability of known data.

Specifically, the analysis was performed by drawing, randomly, from the set of known data, xray scores for the missing data drug or placebo patient cohorts. Thus, on average, the variance for the missing data cohort will approximate that for the cohort with xrays. Then, keeping that particular variance for each missing data cohort fixed, the value of the mean change from baseline was iteratively varied, testing the primary analysis each time to see if a $p<0.05$ conclusion still held. For example, assume the drug patients with xrays showed a mean change from baseline of -2.0 and placebo -6.0 -- that is, drug patients deteriorated one third as much as placebo patients. So, in this scenerio, the sensitivity analysis would incrementally decrease the mean below -2.0 for missing xray drug patients and incrementally increase the mean above -6.0 value for missing xray placebo patient, until that point where the the calculated p value for the overall primary analysis became greater than 0.05. For example, assume this process led to the conclusion that drug patients could deteriorate by, say, -4.5 and placebo patients by -2.5 before the inference was nullified. Here then, the inference would hold until that point where the treatment effect was reversed to a degree of 2.0 -- placebo patients doing better than drug patients. Thus, in this example, the "known data drug effect" would be 4.0 and the "missing data drug effect" would be -2.0, so the overall amount of permitted deviation would be 6.0 -- the maximal amount by which the missing data can deviate from the known data and still have the inference hold.

One can then repeat the process many times, each time using a new, randomly selected, data set (drawn from the known database). In this way one could eventually determine the mean and 95% confidence interval for the maximal amount. If this mean is found to be large and its confidence interval narrow compared to the observed drug effect, then one would usually be reassured that the inference remains valid despite xray dropouts.

An objection to the above sensitivity analysis would occur if the "same variability" assumption was challenged -- ie. if the variability in the missing data cohorts were markedly different from the variability in the known data cohorts. However, the larger the maximal difference found with this analysis, the greater the variability difference that would be needed to undermine the conclusion. In this was of thinking one could fix the

mean difference of the missing data as equal to the mean difference of the known data, and calculate what "deviation of variability" would be needed to undermine the inference. This was not done because of the inherent greater difficulty of thinking in terms of quantities of variability versus quantities of means.

DEMOGRAPHICS / BASELINE COMPARISONS

These data are displayed for all three pivotal trials jointly.

Table 3 Demographic Characteristics

	Age			Gender		Race	
Study and Treatment Group	Mean (years)	<65 yrs N (%)	≥65 yrs N (%)	Male N (%)	Female N (%)	Caucasian N (%)	Other Race N (%)
Pivotal Studies							
US301 (12 months)							
Leflunomide	54.2	141 (78%)	41 (23%)	50 (27%)	132 (73%)	161 (89%)	21 (12%)
Placebo	54.6	97 (82%)	21 (18%)	35 (30%)	83 (70%)	103 (88%)	14 (12%)
Methotrexate	53.3	147 (81%)	35 (19%)	45 (25%)	137 (75%)	162 (89%)	20 (11%)
MN301 (6 months)							
Leflunomide	58.3	90 (68%)	43 (32%)	32 (24%)	101 (76%)	115 (87%)	18 (14%)
Placebo	58.8	65 (71%)	27 (29%)	23 (25%)	69 (75%)	82 (89%)	10 (11%)
Sulfasalazine	58.9	82 (62%)	51 (38%)	41 (31%)	92 (69%)	124 (93%)	9 (7%)
Placebo-Controlled Phase II Study							
YU203 (6 months)							
Placebo	52.8	92 (90%)	10 (10%)	24 (23%)	78 (77%)	102 (100%)	0 (0%)
5 mg/day	50.3	84 (88%)	11 (12%)	16 (17%)	79 (83%)	95 (100%)	0 (0%)
10 mg/day	51.4	87 (86%)	14 (14%)	14 (14%)	87 (86%)	101 (100%)	0 (0%)
25 mg/day	50.0	89 (86%)	15 (14%)	13 (12%)	91 (88%)	104 (100%)	0 (0%)
Active-Controlled Studies							
MN301/303 (12 mos.)							
Leflunomide	57.8	51 (64%)	29 (36%)	20 (25%)	60 (75%)	69 (86%)	11 (14%)
Sulfasalazine	58.7	46 (61%)	30 (40%)	26 (34%)	50 (66%)	69 (91%)	7 (9%)
MN302 (12 mos.)							
Leflunomide	58.3	347 (70%)	154 (31%)	147 (29%)	354 (71%)	495 (99%)	6 (1%)
Methotrexate	57.8	348 (70%)	150 (30%)	143 (29%)	355 (71%)	491 (99%)	7 (1%)

Table 4 RA Disease History

	Duration of RA				DMARD Use	
Study and Treatment Group	Mean (yrs)	≤2 yrs N (%)	>2-10 yrs N (%)	>10 yrs N (%)	Prior DMARD N (%)	Mean No. DMARDs Used
Pivotal Studies						
US301 (12 months)						
Leflunomide	7.0	71 (39%)	66 (36%)	45 (25%)	101 (56%)	0.8
Placebo	6.9	39 (33%)	53 (45%)	25 (21%)	71 (60%)	0.9
Methotrexate	6.5	73 (40%)	72 (40%)	37 (20%)	102 (56%)	0.9
MN301 (6 months)						
Leflunomide	7.6	50 (38%)	43 (32%)	40 (30%)	80 (60%)	1.2
Placebo	5.7	41 (45%)	29 (32%)	22 (24%)	43 (47%)	0.9
Sulfasalazine	7.4	56 (42%)	40 (30%)	37 (28%)	65 (49%)	1.0
Placebo-Controlled Phase II Study						
YU203 (6 months)						
Placebo	8.7	18 (18%)	46 (45%)	38 (37%)	83 (81%)	1.8
5 mg/day	8.1	11 (12%)	59 (62%)	25 (26%)	80 (84%)	1.9
10 mg/day	8.9	14 (14%)	49 (49%)	38 (38%)	81 (80%)	1.6
25 mg/day	9.1	10(10%)	58 (56%)	35 (34%)	84 (81%)	1.8
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	6.4	33 (41%)	27 (34%)	20 (25%)	49 (61%)	1.2
Sulfasalazine	6.5	32 (42%)	27 (36%)	17 (22%)	37 (49%)	0.9
MN302 (12 mos.)						
Leflunomide	3.7	219 (44%)	270 (54%)	12 (2%)	332 (66%)	1.1
Methotrexate	3.8	215 (43%)	265 (53%)	18 (4%)	333 (67%)	1.1

Table 5 Concomitant RA Medication Use			
Study and Treatment Group	NSAID Use N (%)	Corticosteroid Use N (%)	Both N (%)
Pivotal Studies			
US301 (12 months)			
Leflunomide	137 (75%)	98 (54%)	74 (41%)
Placebo	77 (65%)	65 (55%)	46 (39%)
Methotrexate	127 (70%)	96 (53%)	66 (36%)
MN301 (6 months)			
Leflunomide	114 (86%)	60 (45%)	49 (37%)
Placebo	80 (87%)	41 (45%)	35 (38%)
Sulfasalazine	104 (78%)	61 (46%)	47 (35%)
Placebo-Controlled Phase II Study			
YU203 (6 months)			
Placebo	98 (96%)	42 (41%)	40 (39%)
5 mg/day	90 (95%)	32 (34%)	31 (33%)
10 mg/day	97 (96%)	40 (40%)	40 (40%)
25 mg/day	98 (94%)	45 (43%)	40 (39%)
Active-Controlled Studies			
MN301/303 (12 mos.)			
Leflunomide	68 (85%)	29 (36%)	24 (30%)
Sulfasalazine	56 (74%)	32 (42%)	23 (30%)
MN302 (12 mos.)			
Leflunomide	402 (80%)	358 (72%)	285 (57%)
Methotrexate	431 (87%)	323 (65%)	279 (56%)

REF: Appendix Table 12.1

APPEARS THIS WAY
ON ORIGINAL

TRIAL REPORT - US301

OVERVIEW

This was a one year US, multicenter, randomized, double-blind comparison of the clinical and radiologic efficacy of leflunomide (LEF) 20mg/d, methotrexate (MTX) 7.5-15mg/wk, and placebo (PLC) in 482 RA patients with disease at least 6mo duration, ACR "active", and MTX naive, assigned with a 3:3:2 ratio, and stratified by time from last disease-modifying antirheumatic drug (DMARD) of more than 8weeks or not. Patient demographics and baseline characteristics are shown above, along with the other trial data, to facilitate cross trial comparison. US 301 was thus both a difference design - LEF vs PLC, and an equivalence design - LEF vs MTX, using the agreedupon test that the lower limit of the 95% confidence interval of the difference of the two active arms should exceed minus 10%. This protocol specified that patients not demonstrating an ACR20 response on or after 4 months or with adverse reactions requiring withdrawal could (blindly) change treatment (after 4 or 8 week washouts as, if indicated). The changed treatment would be LEF if the patient had started on MTX or PLC; it would be MTX if started on LEF.

PATIENT ACCOUNTABILITY

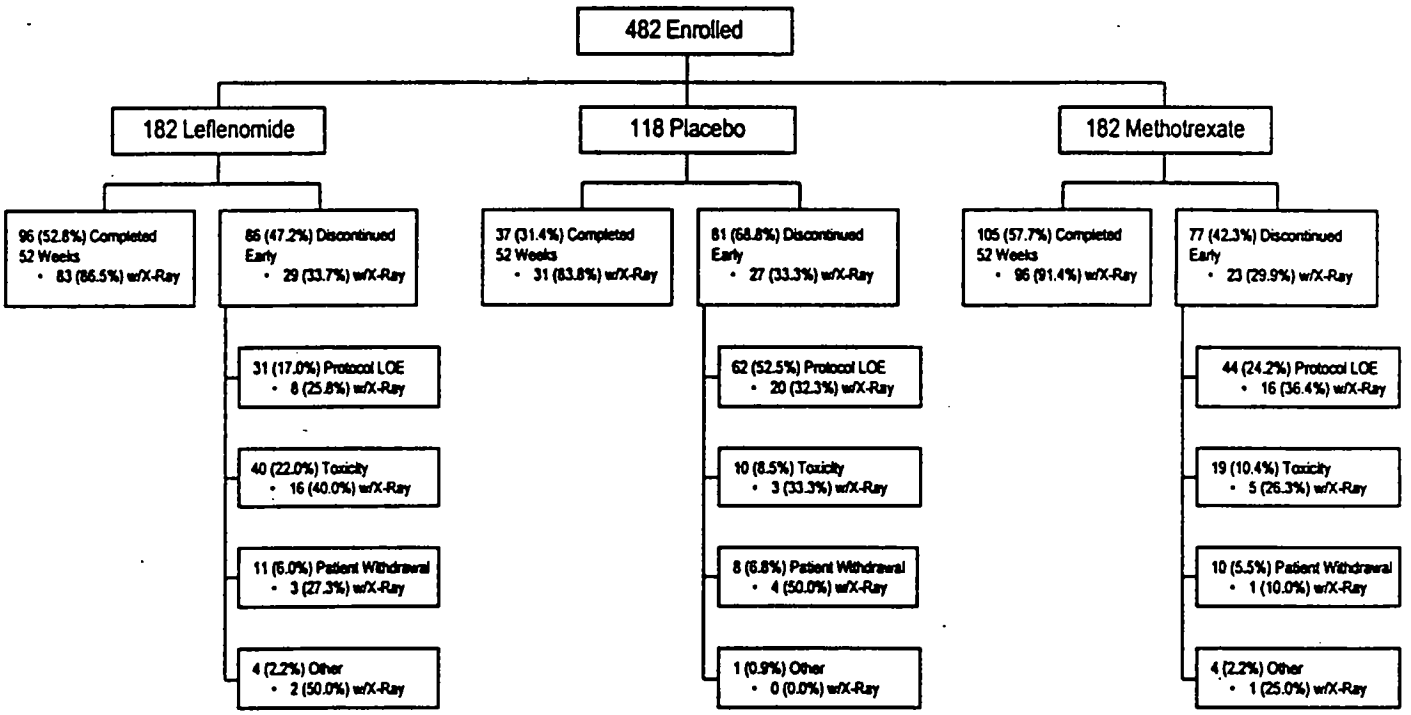
Patients dropping out for defined inefficacy (ACR20 failure at month 4 or beyond), or defined toxicity were eligible to (blindly) continue in the trial on alternative therapy, specifically, switched to LEF if initially on MTX or PLC, and switched to MTX if initially on LEF.

	LEF	MTX	PLC	TOTAL
# randomized	182	182	118	482
# completed	96	105	37	238
# dropped out	86	77	81	244
# eligible for switch	30	42	60	
# electing to switch	24	33	51	

XRAY ACCOUNTABILITY

NDA #20-905

Protocol US301: Flow Chart of Subject and X-Ray Accountability



APPEARS THIS WAY
ON ORIGINAL

EFFICACY ANALYSES

1. PRIMARY CLINICAL: ACR20 RESPONSE AT 12 MONTHS

LEF	74/182 (41%)	LEF v PLC: p=0.0001	LEF v MTX CI=-4.3%, 15.6%
MTX	63/180 (35%)	MTX v PLC: p=0.0001	
PLC	22/118 (19%)		

2. SECONDARY CLINICAL: MEAN AUC UNDER ACR20 RESPONSE CURVE

LEF	23.7wk	LEF v PLC: p=0.0001	LEF v MTX CI=-3.8wk, 6.2wk
MTX	22.7wk	MTX v PLC: p=0.0001	
PLC	12.6wk		

3. PRIMARY RADIOGRAPHIC: SHARP SCORE: TIME ON ALL DRUGS - ALL PATIENTS WITH PAIRED XRAYS

		mean+/- SD		
	n	baseline	change	
LEF	131	23.1+/-34.0	0.5+/-4.5	LEF v PLC: p=0.0007
MTX	138	22.8+/-39.0	0.9+/-3.3	MTX v PLC: p=0.0187
PLC	83	25.4+/-31.3	2.2+/-4.0	LEF v MTX: CI=-2.30,-0.003 (P=0.0494)

4. SECONDARY RADIOGRAPHIC: SHARP SCORE: TIME ON INITIAL DRUG - ALL PATIENTS WITH PAIRED XRAYS FOR BEGINNING / END OF DRUG RX

		mean+/- SD		
	n	baseline	change	
LEF	112	23.22+/-34.86	0.45+/-3.87	LEF v PLC: p=0.0159
MTX	119	22.75+/-39.93	0.82+/-3.08	MTX v PLC: NS
PLC	58	26.43+/-32.66	1.71+/-3.84	LEF v MTX:CI=-2.02, 0.14

5. ADDITIONAL RADIOGRAPHIC ANALYSIS: An analysis of xray data of patients completing the full 12 months therapy on the initially assigned treatment.

		mean	mean	
	n	baseline	change	
LEF	80	24.3	0.6	LEF v PLC: p=0.0761
MTX	95	25.0	1.0	MTX v PLC: NS
PLC	31	28.0	1.6	LEF v MTX:CI=0.0757

DROPOUT ANALYSES

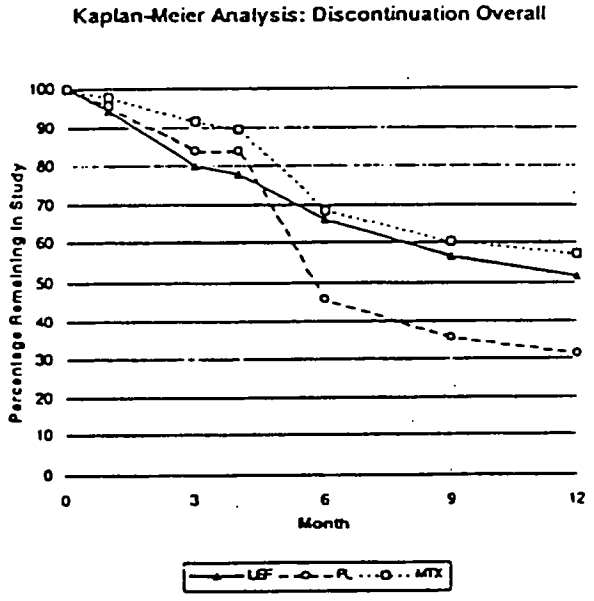
1. MISSING CLINICAL DATA

a. Dropouts by type

Summary of Reasons for Early Withdrawal from the Treatment Phase

Reason for Withdrawal	Leflunomide (N=182)		Placebo (N=118)		Methotrexate (N=182)		Total (N=482)
	N	%	N	%	N	%	
Lack of efficacy	31	17	62	53	44	24	137
Adverse event	40	22	10	9	18	10	68
Lost to follow-up	1	1	0	—	2	1	3
Protocol violation	0	—	1	1	1	1	2
Noncompliance	1	1	0	—	1	1	2
Death	0	—	0	—	1	1	1
Other	13	7	8	7	10	5	31
TOTAL	86	47	81	69	77	42	244

b. Time to dropout



c. Efficacy state at time of dropout

Data showing mean (+/- SD) changes in selected variables (SJC, TJC, MHAQ) for the dropout cohorts, and scatter plots of the same data are in the appendix.

2. MISSING RADIOGRAPHIC DATA

a. Analysis of clinical states of missing xray patients

Study US301

Measure	Subjects with X-ray Data			Subjects without X-ray Data		
	LEF N=130	PLA N=83	MTX N=138	LEF N=49	PLA N=35	MTX N=42
SJC	-6.5	-3.3	-5.6	-3.7	-1.9	-4.5
TJC	-8.2	-3.6	-7.0	-6.4	-1.7	-5.3
MD Global	-3.1	-1.2	-2.5	-2.1	-0.4	-2.0
Pt Global	-2.2	0.1	-1.7	-1.8	0.1	-0.8
Pain	-21.8	-5.3	-19.5	-22.5	-2.4	-9.9
MHAQ	-0.3	0.0	-0.2	-0.3	0.2	0.0
ESR	-7.1	1.4	-8.1	-3.9	5.3	-0.8
CRP	-0.8	0.5	-0.6	-0.2	0.5	-0.2

b. Sensitivity analysis: The sponsor did a sensitivity analysis (described in the section on methodology above) to determine the maximal amount by which the missing xray data could have deviated from the existing data and still have the conclusion hold to a P value of <0.05, The analysis showed that this maximal difference between the mean Sharp scores of the missing LEF and PLC cohorts would have to reach a value of greater than +1.04 in order for the overall analysis to fail to reach statistical significance to the 0.05 level. The 95% confidence interval for this +1.04 figure was found to be (+0.93, +1.15). Since the difference actually found for the patients with xrays was -2.53 (ie. drug patients worsening less than placebo), the sensitivity analysis indicated that the treatment effect would need to be reversed to a degree of approximately 50% before the statistical significance was voided.

CONCLUSION

This trial supplies substantial evidence of a clinical effect and a radiographic effect.

TRIAL REPORT - MN301

OVERVIEW

This was a 6month, multinational, multicenter, double-blind comparison of LEF 20mg/d, sulfasalazine (SSZ) 2gm/d, and PLC in 358 RA patients assigned using a 3:3:2 ratio and stratified on disease duration of less than 2years or not. Patient demographics and baseline characteristics are shown above. An optional 6month extension, continuing the blind, was avaliable for 6month completers (in the NDA this was called trial MN303), and it was elected by 80 of 96 LEF, 76 of 83 SSZ, and 41 of 51 PLC patients, who, per protocol were to be switched to SSZ. Thus, this protocol is an acceptable 12 month equivalence design (LEF vs SSZ), provided (1) the 6 month comparison of the active control (SSZ) validates the assay by demonstrating superiority to PLC, and (2) no major differential effect was introduced in the process of offering the patients the second six month option. MN301 is thus both a difference and an equivalence design. The same allowable maximal small difference of 10% for an equivalence success was used here as in US301. The MN301 protocol required dropout for certain toxicities, and for defined inefficacy after 16 weeks (3 of the following: TJC change of 2 or less, SJC change of 2 or less, no improvement in patient global, no improvement in investigator's global.

The clinical analyses were identical to US301: primary being a comparison of proportions of ACR20 responders at trial completion, and secondary the mean AUC under the ACR20 response curve. X-rays were obtained at baseline, 6 months, and 12 months here; there was no xray at dropout point. Thus, the radiographic analysis was simply a comparison of the mean change in the x-ray Sharp scores from baseline to 6 months, baseline to 12 months, and 6 months to 12 months.

PATIENT ACCOUNTABILITY

	# patients			
	ALL	LEF	SSZ	PLC
randomized	359	133	133	92
completed 6mo rx	230	96	83	51
dropouts	129	37	50	41
elected to continue in 6mo extension	197	80	76	41*

* These patiuents, per protocol, were crossed over to SSZ. Hereafter, they are noted as PLC/SSZ.

	ALL	LEF	SSZ + PLC/SSZ
began 6-12mo trial	197	80	76 + 41
completed 12mo	168	71	68 + 29
dropouts	29	9	8 + 12

XRAY ACCOUNTABILITY

xray "loss": baseline vs 6mo: 89 of 133 LEF patients had xrays
85 of 133 SSZ
62 of 92 PLC

baseline vs 12mo: 91 of 133 LEF
85 of 225 SSZ**

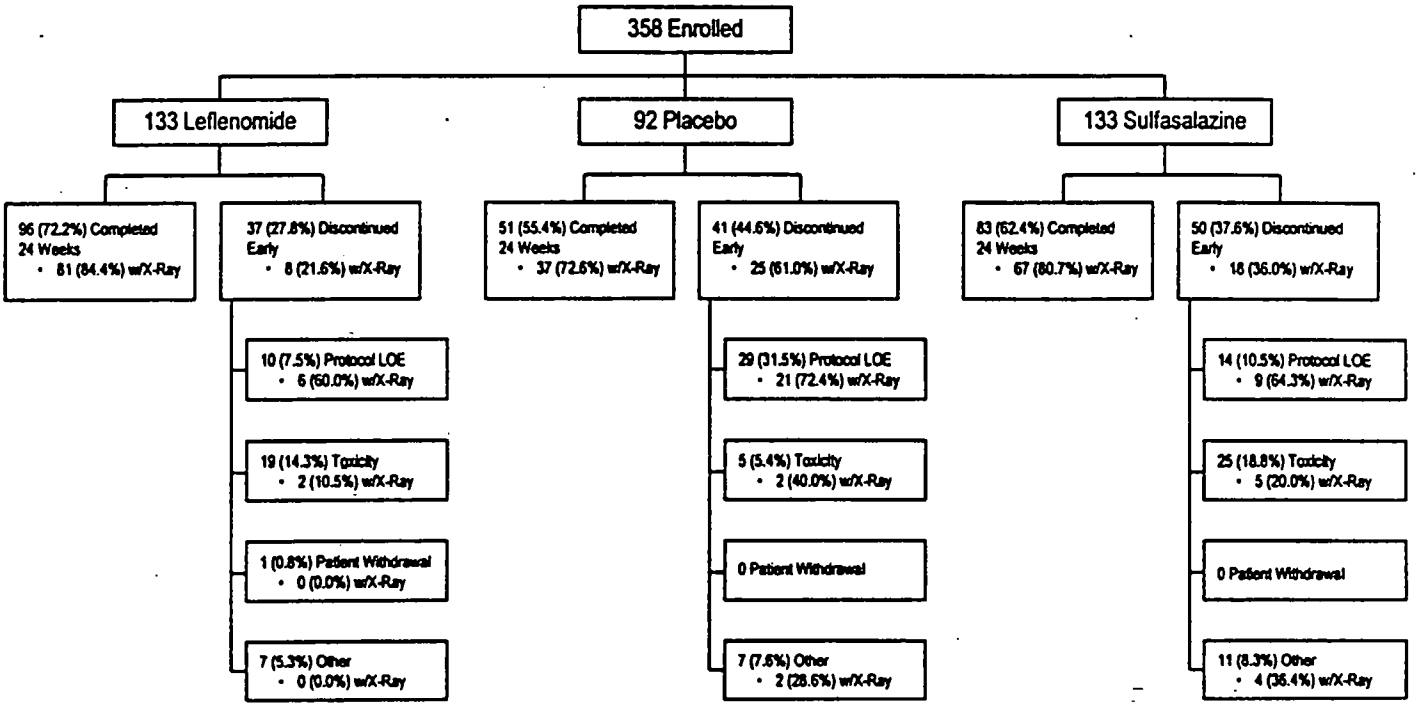
6mo vs 12mo: 60 of 80 LEF
53 of 117 SSZ (117 = 76 SSZ + 41 PLC/SSZ)

** These either had actual SSZ exposure (n = 133 + 41), or potential SSZ exposure (n = 51) but declined or were ineligible.

Below is a flow chart of xray accountability for MN301.

NDA #20-905

Protocol MN301: Flow Chart of Subject and X-Ray Accountability



APPEARS THIS WAY
ON ORIGINAL

EFFICACY ANALYSES

1. PRIMARY CLINICAL: ACR20 RESPONSE AT 6, AND 12 MONTHS

6mo results:		
LEF	63/130 (48.5%)	LEF vs PLC: p = 0.0026
SSZ	59/132 (44.7%)	LEF vs SSZ: CI: -8.3, 15.8
PLC	26/91 (28.6%)	
12mo results:		
LEF	59/130 (45.4%)	LEF vs SSZ: CI: -6.0, 17.9
SSZ	52/132 (39.4%)	

2. SECONDARY CLINICAL: MEAN AUC UNDER ACR20 RESPONSE CURVE

6mo results:		
LEF	11.8wk	LEF vs PLC: p = 0.0001
SSZ	10.5wk	LEF vs SSZ: CI: -0.8wk, 3.6wk
PLC	5.5wk	
12mo results:		
LEF	21.9wk	LEF vs SSZ: CI: -2.6wk, 6.5wk
SSZ	20.1wk	

3. PRIMARY RADIOGRAPHIC: SHARP SCORE: TIME ON ALL DRUGS - ALL PATIENTS WITH PAIRED XRAYS

	n	baseline score	change at 6mo	
0mo vs 6mo				
LEF	89	46.26	-0.06	LEF vs PLC: p = 0.0081
SSZ	85	41.86	1.44	LEF vs SSZ: CI: -5.4, 2.3
PLC	62	46.18	5.60	
0 mo vs 12mo:				
LEF	91	47.19	0.90	LEF vs SSZ: CI: -3.3, 2.5
SSZ	85	41.86	6.44	
6 mo vs 12 mo:				
LEF	60	39.77	0.97	LEF vs SSZ: CI: -2.2, 1.4
SSZ	53	45.70	1.38	

DROPOUT ANALYSES:

1. MISSING CLINICAL DATA

1. Dropouts by type - The table below shows the reasons for withdrawal for the first 6 months of MN301.

Reason for withdrawal from study medication	Number (%) of subjects		
	Leflunomide (N=133)	Placebo (N=92)	Sulfasalazine (N=133)
Lack of efficacy	10 (8)	29 (32)	14 (11)
Adverse events (inc. 1 death)	19 (14)	6 (7)	25 (19)
Refusal/Noncompliance	5 (4)	5 (5)	7 (5)
Other	3 (2)	1 (1)	4 (3)
Total withdrawals	37 (28)	41 (45)	50 (38)

The protocol specified dropout if there was inadequate efficacy at 16 weeks. This was defined as 3 of the following: fall in TJC of 2 or less, fall in SJC of 2 or less, no improvement in patient global, no improvement in investigator's global. Nonetheless, the vast majority of inefficacy dropouts occurred prior to week 16: 8/10 for LEF, 10/14 for SSZ, and 26/29 for PLC.

The table below shows the reasons for withdrawal for the second six months of the trial:

Reason for withdrawal from study medication	Number (%) of subjects		
	Leflunomide (N=80)	Sulfasalazine (N=76)	Plac./Sulfa. (N=41)
Lack of efficacy	4 (5)	2 (3)	2 (5)
Adverse events (inc. 1 death)	2 (3)	5 (7)	9 (22)
Refusal/Noncompliance	2 (3)	0 (0)	1 (2)
Other	1 (1)	1 (1)	0 (0)
Total withdrawals	9 (11)	8 (11)	12 (29)

2. Time to dropout

Summary of Reasons for Early Withdrawal from the Treatment Phase

Reason for Withdrawal	Leflunomide (N=182)		Placebo (N=118)		Methotrexate (N=182)		Total (N=482)
	N	%	N	%	N	%	
Lack of efficacy	31	17	62	53	44	24	137
Adverse event	40	22	10	9	18	10	68
Lost to follow-up	1	1	0	—	2	1	3
Protocol violation	0	—	1	1	1	1	2
Noncompliance	1	1	0	—	1	1	2
Death	0	—	0	—	1	1	1
Other	13	7	8	7	10	5	31
TOTAL	86	47	81	69	77	42	244

3. Efficacy state at time of dropout

Data showing mean (+/-SD) changes in selected variables (SJC, TJC, MHAQ) for the dropout cohorts, and scatter plots of the same are in the appendix.
pp 14-16, 4/21/98 submission

2. MISSING RADIOGRAPHIC DATA

a. Analysis of clinical states of missing xray patients

- Study MN301

Measure	Subjects with X-ray Data			Subjects without X-ray Data		
	LEF N=89	PLA N=62	SSZ N=85	LEF N=41	PLA N=29	SSZ N=47
SJC	-8.1	-3.6	-6.3	-5.2	-2.9	-5.9
TJC	-10.4	-5.0	-8.7	-8.2	-2.8	-6.9
MD Global	-3.0	-1.0	-2.8	-2.0	-0.5	-2.0
Pt Global	-3.0	-1.1	-3.0	-2.2	-0.5	-1.9
Pain	-31.7	-11.1	-22.4	-17.8	-3.9	-15.1
MHAQ	-0.6	-0.1	-0.4	-0.3	0	-0.2
ESR	-6.8	4.9	-20.1	-8.9	0.4	-10.2
CRP	-2.4	0.3	-1.7	-2.0	-0.1	-0.2

Study MN301

Measure	Subjects with X-ray Data			Subjects without X-ray Data		
Subjects Dropped for LOE	LEF N=6	PLA N=21	SSZ N=9	LEF N=4	PLA N=8	SSZ N=5
SJC	-2.3	1.2	-1.9	-1.3	2.5	0.4
TJC	-4.2	0.7	0.2	-5.3	-0.3	3.2
MD Global	-0.4	0.6	-0.6	0.0	1.6	1.5
Pt Global	0.4	0.2	-0.3	-0.6	1.6	1.0
Pain	2.8	10.8	-5.1	-6.3	15.0	-1.0
MHAQ	-0.2	0.2	-0.1	0.0	0.3	0.5
ESR	13.0	13.6	-5.8	-2.3	2.8	29.4
CRP	-0.3	0.0	0.1	-1.5	-1.7	1.4
Subjects Dropped for Safety	N=2	N=2	N=5	N=4	N=8	N=5
SJC	-11.0	2.5	0.0	-5.6	-1.7	-5.5
TJC	-8.5	-3.5	-1.0	-7.0	-3.3	-7.6
MD Global	-1.3	0.0	-1.5	-1.6	-0.8	-2.1
Pt Global	0.0	0.0	-2.0	-2.2	-0.8	-1.6
Pain	8.0	-3.0	-6.6	-17.3	-20.3	-11.7
MHAQ	-0.1	-0.1	0.1	-0.3	-0.2	-0.2
ESR	5.0	-1.0	2.0	-0.8	-4.3	-2.9
CRP	-7.2	-1.7	-3.1	-0.5	0.7	0.4
Subjects Dropped for Other Reasons	N=0	N=2	N=4	N=6	N=4	N=6
SJC		-1	-7	1	-2.8	-4.3
TJC		-3.5	-11	-5.2	-1.3	-5.5
MD Global		-1.3	-1.9	-2.1	-0.6	-1.3
Pt Global		0	-2.5	-1.3	-1.3	-1.3
Pain		-12	-5.8	-11.2	0.3	-8.2
MHAQ		0.3	-0.5	-0.6	-0.2	-0.1
ESR		-0.5	-25.3	-0.5	8.8	-24.3
CRP		-0.7	-1	-2.5	0.5	-0.8

b. Sensitivity analysis: As with US301, the sponsor did a sensitivity analysis (described in the methodology section) to determine the maximal amount by which the missing xray data could have deviated from the existing data and still have the conclusion hold to a P value of <0.05, The analysis showed that this maximal difference between the mean Sharp scores of the missing LEF and PLC cohorts would have to reach a value of greater than +3.50 in order for the overall analysis to fail to reach statistical significance to the 0.05 level. The 95% confidence interval for this +3.50 figure was found to be (+3.08, +3.93). In MN301 the difference actually found for the patients with xrays was -5.17 (ie. drug patients worsening less than placebo), so the sensitivity analysis indicated a deviation of as much as 8.67 could occur before the statistical significance was voided. Thus, in the case of MN301, failure of the inference would also require reversal of the treatment effect, and by an amount slightly more than 50% of the observed effect.

CONCLUSION: This trial also supplies substantial evidence of a clinical and radiographic effect, and the use of placebo for six months adequately validates the assay to ascribe credibility to the 12 month data.

TRIAL MN302

OVERVIEW

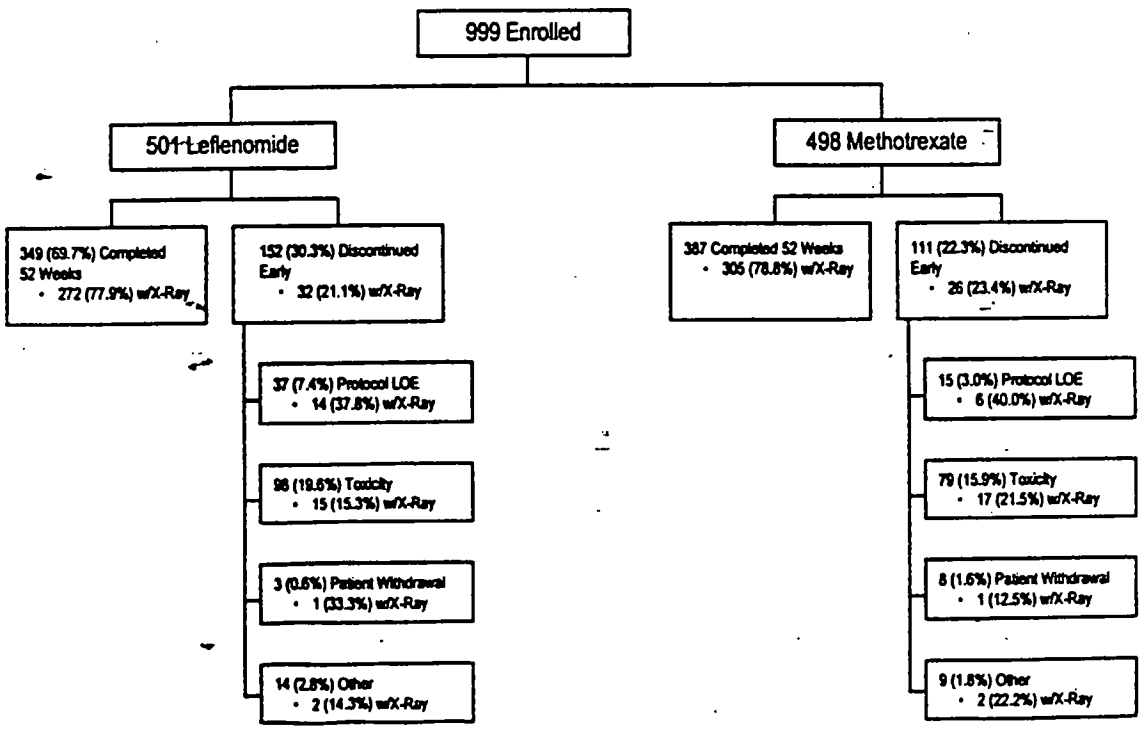
This was a 12 month, multinational, multicenter, double-blind comparison of LEF 20mg/d and MTX 7.5-15mg/wk in 999 RA patients with disease less than 10 yr duration, randomized using a 1:1 ratio of 999 patients. Patient demographics and baseline characteristics are shown earlier. No protocol specified inefficacy dropout criteria were included in the protocol. MN302 was an equivalence design, using the same "equivalence test" as in the above trials (the 95% CI lying fully to the right of -10%). The clinical analyses were identical to trials US301 and MN301; the radiographic analysis as with MN301, because only baseline and 12 month x-rays were obtained. Thus, the primary clinical analysis was a comparison of proportions of ACR20 responding patients at 12 months. All others were failures, and the mean AUC duration (in weeks) under the "ACR20 response curve" was the secondary. The primary radiographic analysis was a comparison of mean change in x-ray Sharp scores from baseline to end-of-trial (regardless of intervening drug changes).

PATIENT ACCOUNTABILITY

	# patients	
	LEF	MTX
randomized	501	498
completed 12 mo	349	387
dropouts	152	111

XRAY ACCOUNTABILITY

Protocol MN302: Flow Chart of Subject and X-Ray Accountability



EFFICACY ANALYSES

1. PRIMARY CLINICAL: ACR20 RESPONDERS AT 12 MONTHS

12mo results:

LEF	215/495	(43.4%)	LEF vs MTX: CI: -19.4, -7.0
MTX	277/489	(56.6%)	(P<0.0001)

2. SECONDARY CLINICAL: MEAN AUC UNDER ACR20 RESPONSE CURVE

12mo results:

LEF	23.0wk	LEF vs MTX: CI: -4.8wk, 0.1wk
MTX	25.4wk	

These results are due, in part, to the earlier onset of LEF compared to MTX, averaging 10.6wk for LEF compared to 14.4wk for MTX.

3. PRIMARY RADIOLOGIC: SHARP SCORE - ALL PATIENTS WITH PAIRED XRAYS

0 mo vs 12mo:

	n	baseline score (=/-SD)	change at 6mo (+/-SD)
LEF	304	24.94+/-31.85	2.19+/-6.65
MTX	331	24.60+/-33.71	1.04+/-14.65

LEF vs MTX: CI: -2.68, 7.96

DROPOUT ANALYSES

1. MISSING CLINICAL DATA

1. Dropouts by type: The table below shows reasons for withdrawal in MN302.

Reason for withdrawal	Number (%) of subjects withdrawn from study medication	
	Leflunomide (N = 501)	Methotrexate (N = 499)
Lack of efficacy	37 (7)	15 (3)
Adverse events	94 (19)	74 (15)
Death	4 (1)	5 (1)
Refusal/noncompliance	11 (2)	14 (3)
Other	6 (1)	3 (<1)
Total	152 (30)	111 (22)

2. Time to dropout - not available for this trial.

3. Efficacy state at time of dropout

Data showing mean (+/-SD) changes in selected variables (SJC, TJC, MHAQ) for the dropout cohorts, and scatter plots of the same data are in the appendix.

2. MISSING RADIOGRAPHIC DATA

a. Analysis of clinical states of missing xray patients

Study MN302

Measure	Subjects with X-ray Data		Subjects without X-ray Data	
	LEF N=304	MTX N=331	LEF N=191	MTX N=158
SJC	-7.6	-9.7	-5.5	-7.5
TJC	-9.7	-10.6	-6.0	-8.0
MD Global	-2.7	-3.3	-1.7	-2.7
Pt Global	-2.7	-3.3	-1.7	-2.5
Pain	-25.0	-32.8	-14.9	-19.4
MHAQ	-0.4	-0.5	-0.3	-0.4
ESR	-12.9	-25.3	-5.8	-15.7
CRP	-2.2	-2.9	-1.3	-1.6

Measure	Subjects with X-ray Data		Subjects without X-ray Data	
Subjects Dropped for LOE	LEF N=14	MTX N=6	LEF N=23	MTX N=9
SJC	2.6	-0.7	-0.5	-2.0
TJC	1.9	-2.5	-1.7	-3.6
MD Global	2.0	0.8	0.8	-0.3
Pt Global	1.1	0.0	0.4	-0.8
Pain	5.0	-13.2	5.6	3.0
MHAQ	0.1	-0.1	0.0	0.1
ESR	6.9	-25.3	-0.8	0.8
CRP	0.9	0.0	0.8	0.2
Subjects Dropped for Safety	N=15	N=17	N=81	N=60
SJC	-4.2	-7.7	-4.8	-6.5
TJC	-7.3	-6.0	-5.3	-6.6
MD Global	-2.2	-1.8	-1.3	-2.1
Pt Global	-2.2	-1.5	-1.6	-2.0
Pain	-21.0	-13.1	-15.2	-9.5
MHAQ	-0.4	-0.2	-0.2	-0.3
ESR	-9.3	-7.7	-5.8	-6.2
CRP	-2.6	-1.6	-1.0	-0.4
Subjects Dropped for Other Reasons	N=3	N=2	N=13	N=11
SJC	-6.7	-1.5	-5.2	-3.9
TJC	-7.7	-4.5	-6.1	-4.6
MD Global	-1.7	-1.3	-1.7	-1.8
Pt Global	-2.5	-1.3	-0.6	-1.1
Pain	-1.3	-12.5	-5.6	-26.9
MHAQ	-0.4	0.3	-0.2	-0.1
ESR	11.3	-13.5	-6.5	0.7
CRP	0.8	-1.9	-0.8	0.4

2. Sensitivity analysis: No sensitivity analysis was done because this trial had no placebo control.

CONCLUSION: This trial showed MTX superior to LEF in the primary clinical endpoint, and equivalent (by the "CI fully to the right of -10%" test) to LEF by the secondary AUC. If US301 is taken to assert that MTX demonstrate radiographic efficacy compared to placebo and the "literature" taken as confirmatory of this assertion, then LEF here is equivalent in radiographic efficacy by the primary radiographic endpoint. However, this conclusion is not necessary to satisfy demonstration of success in "signs and symptoms" and in "retardaton in structural damage".

It would be possible to do a sensitivity analysis here too, in ensure that the radiographic outcome is not a consequence of a bias introduced by missing xrays. The intent could be the same as with US301 and MN301, ie. What degree of deviation from the known can the missing be presumed to show, and still have the inference hold? Assuming this analysis did not undermine the xray conclusion, I would describe the MN302 results as supportive of the clinical and the radiographic claim.

EVIDENCE ON ARTHRITIS-RELATED PHYSICAL FUNCTION

This section is included because of recent interest in "enhanced claims". The two year "prevention of disability" claim (see RA Guidance Document) calls for demonstrating improvement compared to control in a validated physical function measure in RA, such as the Modified Health Assessment Questionaire (MHAQ) and the Patient Elicitation Technique (PET) instruments used here, with no deterioration in a (RA validated) health-related quality-of-life (HR-QOL) instrument such as the SF-36. HR-QOL instruments are health status measures encompassing three domains - physical, social, and psychological -- and they rate patients on a 'wellness -->disability' continuum. The MHAQ was collected in all three trials; the PET and SF-36 were collected only in US301. The instruments are in the attached protocols.

By intention-to-treat analyses the following results were obtained for the MHAQ:

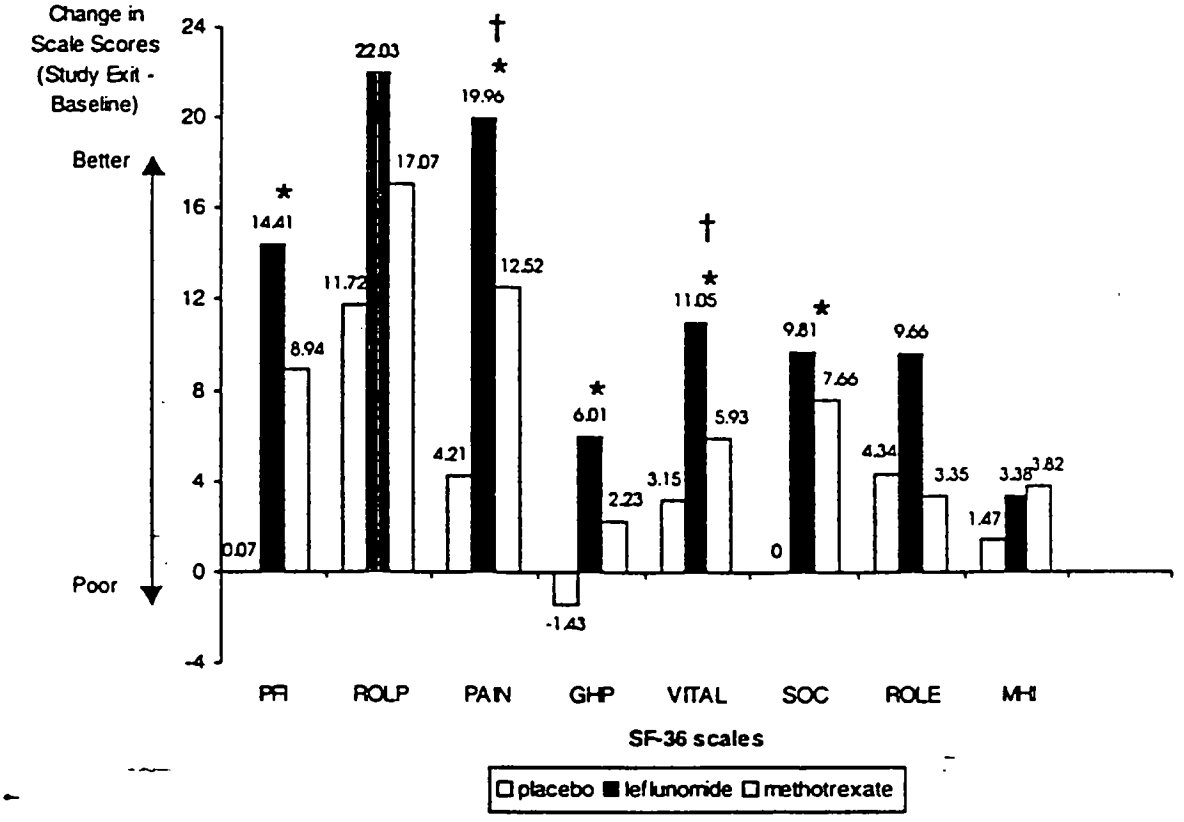
MHAQ:	baseline mn (sd)	change mn (sd)	
US301LEF	0.78 (0.57)	-0.29 (0.54)	LEF vs PLC: p<0.0001
MTX	0.79 (0.50)	-0.15 (0.49)	LEF vs MTX: CI: -0.30, -0.06
PLC	0.87 (0.51)	0.07 (0.50)	(p=0.0027)
MN301LEF	1.14 (0.62)	-0.50 (0.53)	LEF vs PLC: p<0.0001
SSZ	0.98 (0.55)	-0.29 (0.46)	LEF vs SSZ: CI: -0.28, -0.04
PLC	1.09 (0.62)	-0.04 (0.49)	(p=0.0088)
MN302LEF	1.08 (0.63)	-0.37 (0.53)	LEF vs MTX: CI: 0.02, 0.13
MTX	1.06 0.60)	-0.44 (0.50)	(p=0.0118)

LEF is statistically significant better than PLC, MTX, and SSZ in the PLC controlled trials; it is statistically significantly worse than MTX in MN302.

The PET instrument ranks physical activities according to each patient's preferences. The data are "customized", making this arguably the optimal assessment instrument in RA. The weighted top 5 score of the PET has been validated in RA (see: Buchbinder, A&R (1995) 38:11,1568-80; Bombardier, Scand J Rheum (1992) 21:95,29-33). These data for US 301 are shown below:

	PET:baseline mn (sd)	change mn (sd)	
US301LEF	21.2 (10.39)	-6.9 (9.87)	LEF vs PLC: p=0.0001
MTX	20.4 (9.7)	-3.4 (9.90)	LEF vs MTX: CI: -6.42,-1.85
PLC	22.4 (10.36)	-0.7 (8.35)	(p=0.0004)

Health-related Quality-of-Life was assessed using the SF36. The SF36 consists of eight subscales, the results of which are shown graphically below:



* = leflunomide is significantly better than placebo at 0.05 level of significance
† = leflunomide is significantly better than methotrexate at 0.05 level of significance
Note: Number of subjects varies between scales

Scale abbreviations:
PFI = Physical Functioning
ROLP = Role Physical
PAIN = Bodily Pain
GHP = General Health Perception
VITAL = Vitality
SOC = Social Functioning
ROLE = Role Emotional
MHI = Mental Health
REF: US301 Study Report p.87

In five subscales LEF was statistically superior to PLC, and in two it was also statistically superior to MTX.

Finally, the correlation of these physical function and HR-QOL measures were looked at as a function of whether the patient had responded by the traditional clinical ACR20 responder test. These results are noted in the table below.

Table 28 Mean Change in SF-36, HAQ, and PET by ACR Responder Status: US301				
US301 Treatment Group	HAQ Disability Index	PET Top Weighted 5 Score	SF-36 Physical Component	SF-36 Mental Component
Leflunomide				
Responder	-0.7	-10.1	12.4	3.2
Non-Responder	-0.2	-3.2	2.0	-0.3
Placebo				
Responder	-0.4	-5.5	6.6	4.5
Non-Responder	0.1	1.1	-1.3	-0.7
Methotrexate				
Responder	-0.5	-7.9	9.8	3.0
Non-Responder	-0.0	0.5	0.2	-0.8

REF: Appendix Table 31

CONCLUSION: These data are only relevant insofar as they are deemed to have face validity for labelling to help inform clinicians.

APPEARS THIS WAY
ON ORIGINAL

SELECTED ADDITIONAL CLINICAL DATA

The following are shown in the tables below: ACR50, ACR 70, remission, the ACR20 components, AM gel, RF, Hb, and albumin.

Table 8 ACR 50%, ACR 70%, and Treatment Dependent Remission						
Study and Treatment Group	ACR 50% n/N*	%	ACR 70% n/N*	%	% Subjects Achieving Remission** n/N*	%
PIVOTAL STUDIES						
US301 (12 months)						
Leflunomide	61/178	34.3	36/178	20.2	9/178	5.1
Placebo	9/118	7.6	5/118	4.2	1/118	0.8
Methotrexate	41/180	22.8	17/180	9.4	7/180	3.9
MN301 (6 months)						
Leflunomide	43/130	33.1	13/130	10.0	0/130	0.0
Placebo	13/91	14.3	2/91	2.2	0/91	0.0
Sulfasalazine	40/132	30.3	10/132	7.6	2/132	1.5
SUPPORTIVE STUDIES						
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	50/130	38.5	16/130	12.3	1/130	0.8
Sulfasalazine	43/132	32.6	16/132	12.1	2/132	1.5
MN302 (12 mos.)						
Leflunomide	154/495	31.1	49/495	9.9	6/495	1.2
Methotrexate	214/489	43.8	80/489	16.4	21/489	4.3

REF: Appendix Table 408

*N is the number of ITT patients for whom adequate data were available to calculate the indicated rates.

**Remission defined according to ACR criteria of no swollen or tender joints, morning stiffness ≤15 minutes, ESR ≤20 mm/hr for men or ≤30 for women.

APPEARS THIS WAY
ON ORIGINAL

Table 12 Mean Change from Baseline to Endpoint in TJC and SJC (28 joints)						
Study and Treatment Group	N	Mean Baseline TJC	Mean Change at Endpoint	N	Mean Baseline SJC	Mean Change at Endpoint
PIVOTAL STUDIES						
US301 (12 months)						
Leflunomide	179	15.5	-7.7*	179	13.7	-5.7*
Placebo	118	16.5	-3.0*	118	14.8	-2.9*
Methotrexate	180	15.8	-6.6	180	13.0	-5.4
MN301 (6 months)						
Leflunomide	130	18.8	-9.7*	130	16.2	-7.2*
Placebo	91	16.3	-4.3*	91	15.8	-3.4*
Sulfasalazine	132	16.7	-8.1	132	15.3	-6.2
SUPPORTIVE STUDIES						
Placebo-Controlled Phase II Study						
YU203						
Placebo	102	19.9	-4.4*	102	15.9	-3.6*
5 mg/day	95	19.1	-5.2	95	15.4	-4.4
10 mg/day	100	20.1	-7.9*	100	16.1	-6.1*
25 mg/day	101	19.6	-9.5*	101	15.6	-7.3*
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	130	18.8	-9.6	130	16.2	-7.3
Sulfasalazine	132	16.7	-7.9	132	15.3	-6.7
MN302						
Leflunomide	495	17.2	-8.3**	495	15.8	-6.8**
Methotrexate	489	17.7	-9.7**	489	16.5	-9.0**

REF: Appendix Tables 25 and 26

*Indicates statistically significant differences between leflunomide and placebo ($p \leq 0.01$)

**Indicates leflunomide and the active comparator were not statistically equivalent (95% CI not overlapping 0)

Table 14 Mean Change from Baseline to Endpoint in Pain Assessment (in cm from 0-10)			
Study and Treatment Group	N	Mean Baseline Pain Assessment	Mean Change at Endpoint
PIVOTAL STUDIES			
US301(12 months)			
Leflunomide	178	5.9	-2.2*
Placebo	118	6.4	-0.5*
Methotrexate	179	5.8	-1.7
MN301 (6 months)			
Leflunomide	130	6.3	-2.7*
Placebo	91	5.9	-0.9*
Sulfasalazine	132	5.5	-2.0
SUPPORTIVE STUDIES			
Placebo-Controlled Phase II Study			
YU203			
Placebo	101	5.6	-0.7*
5 mg/day	94	5.3	-0.9
10 mg/day	97	5.8	-2.3*
25 mg/day	99	5.5	-2.4*
Active-Controlled Studies			
MN301/303 (12 mos.)			
Leflunomide	130	6.3	-2.8
Sulfasalazine	132	5.5	-1.9
MN302 (12 months)			
Leflunomide	495	5.7	-2.1**
Methotrexate	488	5.9	-2.9**

REF: Appendix Table 29

*Indicates statistically significant differences between leflunomide and placebo ($p \leq 0.0001$)

**Indicates leflunomide and the active comparator were not statistically equivalent (95% CI not overlapping 0)

Table 13 Mean Change from Baseline to Endpoint in Global Assessments (in cm from 0 to 10)

Study and Treatment Group	N	Mean Baseline Patient Assessment	Mean Change at Endpoint	N	Mean Baseline Physician Assessment	Mean Change at Endpoint
PIVOTAL STUDIES						
US301						
Leflunomide	178	5.6	-2.1*	179	6.1	-2.8*
Placebo	118	5.8	0.1*	118	6.2	-1.0*
Methotrexate	179	5.4	-1.5	179	5.9	-2.4
MN301 (6 months)						
Leflunomide	130	6.7	-2.8*	130	6.6	-2.7*
Placebo	91	6.4	-0.9*	91	6.2	-0.8*
Sulfasalazine	132	6.6	-2.6	132	6.3	-2.5
SUPPORTIVE STUDIES						
Placebo-Controlled Phase II Study						
YU203						
Placebo	102	6.4	-1.3*	102	6.1	-1.4*
5 mg/day	95	6.0	-1.5	95	5.8	-1.8
10 mg/day	99	6.1	-2.7*	100	6.0	-2.7*
25 mg/day	101	5.8	-2.6*	101	5.8	-2.7*
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	130	6.7	-2.7	130	6.6	-2.7
Sulfasalazine	132	6.6	-2.5	132	6.3	-2.5
MN302						
Leflunomide	495	6.4	-2.3**	495	6.3	-2.3**
Methotrexate	489	6.6	-3.0**	489	6.4	-3.1**

REF: Appendix Tables 27 and 28

Table 15 Mean Change from Baseline to Endpoint in Total MHAQ/HAQ Score			
Study and Treatment Group	N	Mean Baseline MHAQ (or HAQ) *	Mean Change at Endpoint
PIVOTAL STUDIES			
US301 (12 months)			
Leflunomide	178	0.78	-0.29**,***
Placebo	118	0.87	0.07**
Methotrexate	179	0.79	-0.15***
MN301 (6 months)			
Leflunomide	116	1.14	-0.50**,***
Placebo	81	1.09	-0.04**
Sulfasalazine	113	0.98	-0.29***
SUPPORTIVE STUDIES			
Placebo-Controlled Phase II Study			
YU203			
Placebo	100	5.6	-0.31**
5 mg/day	93	1.35	-0.24
10 mg/day	96	1.50	-0.60**
25 mg/day	99	1.35	-0.55**
Active-Controlled Studies			
MN301/303 (12 mos.)			
Leflunomide	116	1.1	-0.49***
Sulfasalazine	113	.98	-0.28***
MN302 (12 months)			
Leflunomide	477	1.08	-0.37***
Methotrexate	470	1.06	-0.44***

REF: Appendix Table 30

* HAQ was used in MN203; MHAQ was used in Phase III studies

**Indicates statistically significant differences between leflunomide and placebo ($p \leq 0.0001$)

***Indicates leflunomide and the active comparator were not statistically equivalent ($p \leq 0.05$)

Table 16 Mean Change from Baseline to Endpoint in ESR and CRP						
Study and Treatment Group	N	Mean Baseline ESR (mm/Hr)	Mean Change at Endpoint	N	Mean Baseline CRP (mg/dl)	Mean Change at Endpoint
PIVOTAL STUDIES						
US301 (12 months)						
Leflunomide	162	38.4	-6.26*	168	2.1	-0.62*
Placebo	111	37.2	2.56*	115	2.4	0.47*
Methotrexate	163	33.9	-6.48	172	1.9	-0.50
MN301 (6 months)						
Leflunomide	130	55.6	-7.48*****	130	4.5	-2.26*
Placebo	91	52.3	3.44*	91	4.1	0.16*
Sulfasalazine	132	50.5	-16.56**	130	3.4	-1.19
SUPPORTIVE STUDIES						
Placebo-Controlled Phase II Study						
YU203 (6 months)						
Placebo	99	48.1	3.09*	96	2.2	0.81*
5 mg/day	94	50.4	4.22	89	1.9	0.50
10 mg/day	99	56.0	-4.93*	95	2.5	-0.24
25 mg/day	100	45.4	-5.32*	97	1.4	-0.38*
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	130	55.6	-8.44**	128	4.6	-2.29**
Sulfasalazine	132	50.5	-13.82**	128	3.4	-0.79**
MN302 (12 months)						
Leflunomide	494	51.0	-10.12**	495	4.2	-1.86**
Methotrexate	488	51.6	-22.18**	489	4.1	-2.45**

REF: Appendix Tables 32 and 33

Table 17 Mean Change from Baseline to Endpoint in Morning Stiffness and RF						
Study and Treatment Group	N	Mean Baseline Morning Stiffness (mins.)	Mean Change at Endpoint	N	Mean Baseline RF (IU/mg)	Mean Change at Endpoint
PIVOTAL STUDIES						
US301 (12 months)						
Leflunomide	177	202.4	-101.4*	64	304.2	-149.7*
Placebo	117	164.7	14.7*	60	268.1	33.2*
Methotrexate	175	226.2	-88.7	59	200.8	-44.7
MN301 (6 months)						
Leflunomide	130	142.5	-93.0*	124	348.9	-141.1*
Placebo	91	98.1	-6.8*	90	330.6	17.5*
Sulfasalazine	132	110.0	-42.4	125	368.9	-154.1
SUPPORTIVE STUDIES						
Placebo-Controlled Phase II Study						
YU203 (6 months)						
Placebo	102	113.6	-33.7*	102	346.5	-50.2
5 mg/day	95	111.3	-48.3	95	291.6	14.3
10 mg/day	100	91.1	-55.3*	98	332.2	-72.9
25 mg/day	101	94.4	-71.8*	100	239.6	-41.6
Active-Controlled Studies						
MN301/303 (12 mos.)						
Leflunomide	130	142.5	-85.3	122	348.5	-124.2
Sulfasalazine	132	110.0	-35.1	123	372.1	-107.8
MN302 (12 months)						
Leflunomide	495	139.7	-63.7**	495	284.8	-96.1
Methotrexate	489	135.9	-86.6**	489	294.2	-116.7

REF: Appendix Tables 34 and 35

*Indicates statistically significant differences between leflunomide and placebo ($p \leq 0.01$)

**Indicates leflunomide and the active comparator were not statistically equivalent (95% CI not overlapping 0)

Table 18 Summary of Normalized Laboratory Values in ACR Responders*					
Study and Treatment Group	Normalized ESR n/N	Normalized CRP n/N	Normalized Platelet Count n/N	Normalized Hemoglobin n/N	Normalized Albumin n/N
PIVOTAL STUDIES					
US301 (12 mos.)					
Leflunomide	21/31	18/23	16/28	10/13	6/10
Placebo	4/9	1/4	2/6	1/4	1/2
Methotrexate	17/32	17/26	9/14	3/12	6/9
MN301 (6 mos.)					
Leflunomide	19/31	26/44	13/22	13/20	10/17
Placebo	6/13	5/11	1/3	2/3	2/4
Sulfasalazine	39/54	24/33	14/25	3/5	9/18
Placebo-Controlled Phase II Study					
YU203 (6 mos)					
Placebo	7/14	1/8	0/3	3/11	–
5 mg/day	1/4	3/6	1/1	3/7	–
10 mg/day	9/11	12/22	5/10	11/17	–
25 mg/day	14/18	10/18	2/6	8/10	–
SUPPORTIVE STUDIES					
Active-Controlled Studies					
MN301/303 (12 mos.)					
Leflunomide -	20/25	22/30	9/15	11/15	11/13
Sulfasalazine -	25/34	21/26	15/21	6/9	10/13
MN302 (12 mos.)					
Leflunomide	110/154	109/173	70/110	36/55	44/67
Methotrexate	192/261	167/222	90/124	56/68	75/99

REF: Appendix Table 370

*n=number of ACR responders with normalized lab value

N=total number with normalized lab value

CONCLUSION: These data are relevant to labelling discussions.

SELECTED BY-PATIENT RADIOGRAPHIC DATA

The tables below show the percentages of patients who showed radiographic progression, defined either by having a Sharp score change of more than three, or by showing newly eroded joints.

Table 21 Percentage of Subjects with Disease Progression (Using Sharp Erosion Scores for All Studies)

Study and Treatment Group	N	% of Subjects with Disease Progression		Subjects with Newly Eroded Joints	
		N	%	N	%
PIVOTAL STUDIES					
US301 (12 months)					
Leflunomide	131	4	3.1*	17	13.0
Placebo	83	10	12.0*	18	21.7
Methotrexate	138	6	4.3	20	14.5
MN301 (6 months)					
Leflunomide	89	3	3.4*	14	15.7
Placebo	62	10	16.1*	17	27.4
Sulfasalazine	85	4	4.7	18	21.2
SUPPORTIVE STUDIES					
Active-Controlled Studies					
MN301/303 (12 mos.)					
Leflunomide	92	8	8.8	16	17.6
Sulfasalazine	85	6	7.1	20	23.5
MN302					
Leflunomide	304	33	10.9	75	24.7
Methotrexate	331	34	10.3	70	21.2

REF: Appendix Tables 42, 371 and 372

*Indicates statistically significant differences between leflunomide and placebo ($p \leq 0.05$)

**Indicates leflunomide and the active comparator were not statistically equivalent (95% CI not overlapping 0)

CONCLUSION: These data show other ways of analyzing xray progression in RA.

EXPLORATORY ANALYSES: CORRELATION OF CLINICAL & RADIOGRAPHIC

The following table shows the intention-to-treat results of an analysis of total Sharp score changes from baseline in patients with available xrays in the three trials.

Table 41 **Hand and feet X-ray: Total Sharp score for ACR responders vs non-responders in controlled studies – intention-to-treat population**

Study	ACR response	Time point[a] (LOCF)	Leflunomide				Placebo			Active comparator[b]			Lefl. vs placebo		Lefl. vs active comp.	
			Daily dose	N	Mean	SD	N	Mean	SD	N	Mean	SD	p-value	95% CI	p-value	95% CI
US301	No resp.	Baseline	20 mg	58	20.24	25.04	56	26.73	32.88	72	18.47	33.63	-	-	-	-
		Change at endpoint		58	0.98	2.99	56	2.70	4.04	72	1.35	3.77	0.1578	[-2.77,0.46]	0.3498	[-2.33,0.83]
	Resp.	Baseline		71	25.45	40.26	27	22.56	28.12	66	27.44	43.90	-	-	-	-
		Change at endpoint		71	0.20	5.51	27	1.04	3.56	66	0.36	2.49	0.0892	[-6.56,0.48]	0.1381	[-3.14,0.44]
MN301	No resp.	Baseline	20 mg	39	48.87	54.25	44	44.91	52.94	35	43.26	52.73	-	-	-	-
		Change at endpoint		39	-1.74	18.41	44	5.98	10.68	35	-0.43	12.97	0.0532	[-13.4,0.10]	0.8345	[-8.95,7.25]
	Resp.	Baseline		50	44.22	52.37	18	49.28	53.58	50	40.88	52.04	-	-	-	-
		Change at endpoint		50	1.26	2.48	18	4.67	7.55	50	2.74	12.90	0.0164	[-5.69,-0.60]	0.4528	[-5.27,2.37]
MN301/303	No resp.	Baseline	20 mg	41	51.76	57.01	NA	NA	NA	37	48.89	56.23	-	-	-	-
		Change at endpoint		41	0.80	6.41	NA	NA	NA	37	-0.51	12.62	NA	NA	0.0863	[-0.68,9.98]
	Resp.	Baseline		50	43.44	49.21	NA	NA	NA	48	36.44	48.44	-	-	-	-
		Change at endpoint		50	0.98	4.27	NA	NA	NA	48	2.98	13.17	NA	NA	0.5220	[-6.01,3.07]
MN302	No resp.	Baseline	20 mg	123	29.58	33.78	NA	NA	NA	99	24.43	23.43	-	-	-	-
		Change at endpoint		123	2.98	6.37	NA	NA	NA	99	3.79	10.99	NA	NA	0.2064	[-6.60,1.48]
	Resp	Baseline		181	21.78	30.16	NA	NA	NA	232	24.67	37.29	-	-	-	-
		Change at endpoint		181	1.66	6.80	NA	NA	NA	232	-0.13	15.72	NA	NA	0.2602	[-1.85,6.84]
MN303	No resp.	Baseline	20 mg	21	38.33	49.11	NA	NA	NA	14	70.79	55.51	-	-	-	-
		Change at endpoint		21	0.67	8.13	NA	NA	NA	14	2.00	4.11	NA	NA	0.4321	[-11.0,4.90]
	Resp.	Baseline		39	40.54	44.38	NA	NA	NA	39	36.69	50.19	-	-	-	-
		Change at endpoint		39	1.13	4.80	NA	NA	NA	39	1.15	2.32	NA	NA	0.4630	[-1.28,2.78]

Note: Hand and feet X-rays were not done in Study YU203.

p-values and 95% CIs are based on adjusted means. If the subject's assessment was missing at the given time point, the last observation was carried forward (LOCF)

The Sharp score method is described in Sharp, J.T. Scoring Radiographic Abnormalities in Rheumatoid Arthritis. Radiologic Clinics of North America, March, 1996, Vol 34:2, pp 233-241, and is based on assessment of joint space narrowing and erosions of the hands, wrists and feet. X-rays for subjects in Studies US301, MN301, MN302, and MN303 were evaluated by Prof. John T. Sharp, Georgia, USA.

[a] Change = value at indicated timepoint minus value at baseline

[b] Active comparator: sulfasalazine in Studies MN301 and MN303, methotrexate in Studies MN302 and US301. Subjects who received placebo in Study MN301 switched to sulfasalazine in the extension study MN303.

Comment: This table, to my mind, defies any overall, coherent interpretation, a fact which could have many different explanations. The correlation of the clinical and radiographic responses in RA is obviously of immense interest to clinicians, and this database, to date, is by far the largest. The absence of any good understanding of any such correlation underlay the agency's decision to make the radiographic claim in RA contingent on clinical success (see RA Guidance Document). Thus, these data, in themselves, are of tremendous interest to the clinical research community, and for this reason these data may, despite the absense of any conclusion, or maybe even because of because no clear conclusion, also merit includion in the label.

LEFLUNOMIDE / METHOTREXATE COADMINISTRATION

A pilot study (Trial F01) was conducted on the open use of both LEF and MTX concomitantly. Drs. Michael Weinblatt and Joel Kremer were the primary investigators. They enrolled 30 RA patients with average disease duration of 13.6 yr, average number of failed DMARDs of 2.9, and average tender/swollen joint counts of 17/16, to open administration for 6 months of LEF at 20mg/d. 53% of these patients showed a clinical response by ACR20. A 12 patient PK substudy showed no significant drug-drug interaction with no changes in the clearance or AUCs of either drug. A randomized controlled trial of LEF versus PLC in patients on background MTX is planned.

Seventeen patients in this study demonstrated AST (SGOT) elevations of >1.2 ULN. Two were discontinued from the study, both with >3x increases. Of the remaining 15 continuing on LEF and MTX, 12 reversed to ≤1.2X and 3 remained >1.2x ULN. The outcome of these patients is shown in the table below:

Table 41 AST (SGOT) Elevations> 1.2xULN in Study F01

	Number of Subjects (% out of 30)
Number in Cohort	17 (57)
Patients with NSAID usage	11 (37)
Discontinued LEF due to LFT/low albumin	2 (7)
Remained ≤ 2xULN on Study Rx.	11 (37)
Reversed to ≤ 1.2xULN at Endpoint	12 (40)
Dose reduction	1 (3)
No Dose reduction	11 (37)
Reversed to ≤ 1.2xULN after LEF disc.	2 (7)
Total Reversed to ≤ 1.2xULN	14 (47)
> 1.2xULN at Endpoint	3 (10)
On NSAIDs	3 (10)

Further analysis of AST elevations more relevant to labelling revealed 5/30 patients with 2-3 fold increases in AST (2 resolved, ie. <1.2X, despite continuation; 3 resolved with discontinuation), and 5/30 patients with a >3 fold increase in AST (all of which with discontinuation).

Three patients in this trial met "ACR criteria" for liver biospy used for methotrexate treatment. These criteria call for either 5 of 9 AST elevations (of any degree) or six or more consecutive monthly elevations of AST (to any degree). The histology showed Roegnik Grade I in one and Roegnik Grade IIIA in two (Grades IIIB and IV are indications to permanently discontinue MTX). These patients are described in the safety review below.

CONCLUSION: This study, and the LFT experience,should be detailed in the label. No efficacy inference is possible.