

# Informing Response Criteria for Psoriatic Arthritis. I: Discrimination Models Based on Data from 3 Anti-Tumor Necrosis Factor Randomized Studies

DAFNA D. GLADMAN, BRIAN D.M. TOM, PHILIP J. MEASE, and VERNON T. FAREWELL

**ABSTRACT. Objective.** To develop statistical models, based on the analysis of data from phase III randomized placebo-controlled trials of tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) inhibitors over a 24-week period, that may inform the definition of response measures for clinical trials in psoriatic arthritis (PsA).

**Methods.** Data from phase III randomized controlled trials with anti-TNF agents were used. A training set using baseline and 24-week data from 2 trials was used to derive the models, which were then tested on a dataset using baseline and interim data from the third trial, and baseline and interim data from the first 2 trials. Logistic regression, tree analysis, and factor analysis were considered in the development of the models. Receiver-operating characteristic curves were constructed and area under the curve (AUC) calculated to assess performance of the models.

**Results.** Two models were derived. One was based on differences between baseline and last-visit values, which identified the current 68 tender joint count (TJC68), baseline and change in C-reactive protein (CRP), and the measure with the highest difference among the patient and physician global assessment of disease activity (GDA), patient assessment of pain and the Health Assessment Questionnaire (HAQ). The second model was based on percentage change from baseline and included TJC68, CRP, physician GDA, patient global assessment of arthritis pain, and HAQ. Both models provided high AUC of at least 0.8 for both the training and testing sets.

**Conclusion.** Models for discriminating joint disease response patterns in PsA were derived from data from randomized controlled trials. These models can now be used to inform further consideration of response measures for trials. (First Release July 1 2010; J Rheumatol 2010;37:1892–7; doi:10.3899/jrheum.091172)

## Key Indexing Terms:

PSORIATIC ARTHRITIS TUMOR NECROSIS FACTOR INHIBITOR RESPONSIVENESS

Psoriatic arthritis (PsA) is an inflammatory arthritis associated with psoriasis that is clearly distinguished from rheumatoid arthritis (RA)<sup>1</sup>. Instruments used in clinical trials in PsA to date include the American College of Rheumatology (ACR) improvement/response criteria and

the disease activity score (DAS) and DAS28, both developed for RA<sup>2,3,4</sup>. The Psoriatic Arthritis Response Criteria (PsARC), a composite instrument originally developed for a sulfasalazine study in PsA, has also been used<sup>5</sup>. None were validated for PsA prior to their use in clinical trials. However, in recent trials, both the ACR 20% improvement criteria (ACR20) and the PsARC demonstrated efficacy of tumor necrosis factor agents (anti-TNF), as well as leflunomide, in patients with PsA<sup>6,7,8,9,10,11</sup>. A recent investigation of responsiveness based on the results of phase II trials with etanercept and infliximab concluded that ACR20 was better than PsARC, but both instruments were useful response measures for the assessment of arthritis in PsA clinical trials<sup>12</sup>. However, the investigation did not attempt to derive any response criteria from the data.

A core set of domains for PsA was introduced by OMER-ACT (Outcome Measures in Rheumatology Clinical Trials), and included assessment of joint disease by 68 tender joint counts (TJC) and 66 swollen joint counts (SJC), skin disease by the Psoriasis Area and Severity Index (PASI), patient global assessment of disease activity (PtGDA), and patient assessment of pain, physical function and health-related quality of life<sup>13</sup>. These features have been included in randomized controlled trials in PsA. Although acute-phase

---

From the University of Toronto; the Psoriatic Arthritis Program, Centre for Prognosis Studies in The Rheumatic Diseases, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada; MRC Biostatistics Unit, Institute of Public Health, Cambridge, England; University of Washington School of Medicine; and Rheumatology Research, Swedish Medical Center, Seattle, Washington, USA.

Supported by the MRC Biostatistics Unit (MRC Grant U.1052.00.009), Cambridge, England; and The Krembil Foundation, Toronto, Ontario, Canada.

D.D. Gladman, MD, FRCPC, Professor of Medicine, University of Toronto, Senior Scientist, Toronto Western Research Institute, Director, Psoriatic Arthritis Program, Centre for Prognosis Studies in The Rheumatic Diseases, University Health Network, Toronto Western Hospital; B.D.M. Tom, PhD, Investigator Scientist; V.T. Farewell, PhD, Senior Scientist, MRC Biostatistics Unit, Institute of Public Health; P.J. Mease, MD, Seattle Rheumatology Associations; Director, Rheumatology Research, Swedish Medical Center; Clinical Professor, University of Washington School of Medicine.

Address correspondence to Dr. D. Gladman, Toronto Western Hospital, 399 Bathurst Street, 1E410B, Toronto, Ontario M5T 2S8, Canada.

E-mail: dafna.gladman@utoronto.ca

Accepted for publication April 30, 2010.

reactants were not included in the core set accepted by OMERACT, primarily because they are elevated in only half the patients with PsA, these variables are included in clinical trials as they are prognostically important in this disease.

We examined data from 3 phase III randomized placebo-controlled trials of anti-TNF agents to determine which items best distinguish drug-treated from placebo-treated patients. These anti-TNF trials were chosen because of the unequivocal effectiveness of these drugs in treating patients with PsA at 24 weeks<sup>8,9,11</sup>. We describe the development of models to discriminate between these patient groups. These models were developed primarily through statistical considerations although clinical considerations were taken into account.

## MATERIALS AND METHODS

In total, 366 patients were randomized to the placebo arms of the trials and 354 to the active treatment arms of the trials. Amgen Inc. provided the data on the etanercept trial, in which 205 patients were randomized to receive placebo (n = 104) or 25 mg etanercept (n = 101) subcutaneously twice weekly for 24 weeks<sup>8</sup>. We use the data on these 205 patients from the baseline, 12-week, and 24-week assessments. PASI scores were not supplied. Centocor Ortho Biotech Inc. supplied the data from the infliximab trial (IMPACT 2), in which 200 patients with active PsA were randomized either to 5 mg/kg infusions of infliximab (n = 100) or to placebo (n = 100) at Weeks 0, 2, 6, 14, and 22. A subset of 160 patients, 77 in the infliximab arm and 83 in the placebo arm, who had baseline PASI information available, was provided and we used data for baseline, 14, and 24 weeks<sup>11</sup>. Abbott Laboratories provided the data from the adalimumab trial (ADEPT), in which 313 patients were randomized to receive placebo (n = 162) or 40 mg adalimumab (n = 151) subcutaneously every other week for 24 weeks<sup>9</sup>. We used the data collected at baseline and at 12 and 24 weeks. PASI was available for 69 patients in each of the arms.

*Assessments and measures available.* A common combined dataset was required across the 3 trials. Inevitably, although the same items were requested from all the companies, the information on these items (if supplied) was not necessarily of the same form/structure or level of detail across trials.

The PtGDA and physician global assessment of disease activity (MDGDA) and the patient assessment of pain were rated using either a Likert scale or a visual analog scale (VAS) or both, depending on the trial. Either a 0–10 or 0–100 VAS or a 6-point Likert scale (0–5) was used to assess patients' arthritis pain in the 3 trials.

These measures were all provided or were calculable from the data provided by the trials: overall Health Assessment Questionnaire (HAQ) score, Medical Outcomes Study Short-form (SF-36) physical and mental components (SF-36PCS and SF-36MCS) and corresponding domains, C-reactive protein (CRP), age at start of the study, ages at onsets of psoriasis and PsA, disease duration at the start of the study, and the patient identifier and treatment indicator. However, information on morning stiffness, dactylitis and enthesitis, and the PASI was not provided by all the trials.

*Derived measures.* For tender and swollen joint counts we chose the 68 TJC and the 66 SJC, to reflect the recommendations of OMERACT<sup>13</sup>. We rescaled the 78 TJC and 76 SJC provided in one trial by factors of 68/78 and 66/76, respectively, as individual joint information was not available to directly calculate the 68 TJC and 66 SJC. More sophisticated regression modeling of the relationship between the joint counts did not further enhance model fitting. To create common variables for arthritis pain, and the PtGDA and MDGDA, we chose to define Likert-type variables for these measures from the VAS scales when provided by the trials.

Differences and percentage change variables were computed for each subject from baseline to the final timepoint and from baseline to the intermediate timepoint for the various core measures in the combined dataset.

*Statistical methods.* Although not a gold standard for responsiveness, the treatment indicator of whether randomized to the placebo or drug arm of a trial was used as the outcome measure (i.e., the dependent variable). Information that discriminates between these groups should reflect measures that would discriminate between responding and nonresponding patients. We split the combined dataset into 2 sets: a training dataset on which we built models, and a test dataset on which we validated the models constructed. We used as our training dataset the joint baseline and 24-week (final timepoint) data from 2 trials. We used as our main test dataset the baseline and intermediate data from the third trial (external validation). Additional validation was performed using the joint baseline and intermediate data from the other 2 trials.

Two separate statistical investigations were conducted. One focused on the differences in the various measures over the 24-week period and one on the percentage change in the various measures over the 24-week followup period.

For univariate analyses, initial comparisons of various change measures were done by calculating the mean and SD for the groups of patients randomized to the placebo and drug arms of the training dataset and by calculating effect size (ES, defined as mean change from baseline within group/SD baseline), group effect size (GES, defined as mean change between groups in difference measure/pooled SD for difference measure), Guyatt's effect size (GUES, defined as the mean in difference measure for the drug group/SD of the difference measure for the placebo group), and standardized response mean (SRM, defined as mean change from baseline/SD change from baseline), and by performing t-tests and fitting "univariate" logistic regression models to this treatment indicator, with and without adjusting/stratifying for trial<sup>14</sup>. Note that it is the linear predictor or discriminant function that is of interest here and not the probability of a particular "outcome" because the fraction of patients in the 2 outcome groups is fixed by design.

To identify multivariately which variables contributed most to discriminating between treatment arms, we used an automatic stepwise procedure, based on the Akaike information criterion, to eliminate variables thought to be not statistically important for discrimination in the logistic regressions, and arrived at candidate models<sup>15</sup>. Further, we checked for internal validity of these candidate models by using cross-validation with 10% random removal of subjects in the training set. In addition, if necessary, we allowed for further refinements of these models by manual intervention based on informed input and the principle of parsimony.

Finally, a third strategy was also considered, based on identifying domains for use in the logistic regression analyses. These domains were informed partly through use of factor analysis, with varimax rotation, on the training dataset to see whether variables would cluster into separate and clinically sensible domains based on the factor loadings, and through the realization that there may be strong correlations among the variables. The factor analyses were performed on the difference measures, with and without inclusion of the corresponding baseline variables. Note that the latent factors derived from these analyses were not considered as explanatory variables in any of the regression models because for any regression analysis, the primary focus should be on the relationship of the explanatory variables and outcome, and not on the distribution of the explanatory variables. However, the variables found to cluster together from these factor analyses were indicative of what variables were attempting to identify the same underlying structure and therefore possibly should not be included together simultaneously as main effects in any analysis on the treatment indicator.

Receiver-operating characteristic (ROC) curves based on the indices developed from these various candidate logistic models were constructed and the area under the ROC curve (AUC) calculated to assess performance in the training dataset. Further, we validated them using the testing sets.

Investigations were performed on the effect of PASI and other variables not available for all trials on the final "domain" model adopted.

## RESULTS

*Univariate results.* Table 1 shows that all measures improved, as indicated by the mean changes, ES, and SRM over the 24-week period from baseline in both the placebo and the drug groups of the 2 trials. The larger improvements were associated with being in the active drug group. There was an indication that the patients' and physicians' measures of disease activity, pain, physical quality of life, and functional limitation, and the laboratory measurement CRP all performed better regarding discriminating between active drug and placebo than did the TJC and SJC measures.

*Multivariate analyses for difference measures.* Initial multivariate logistic regression models focusing on the difference measures produced candidate difference models that did not include joint counts (either swollen or tender). The performances of these candidate models were good, with AUC between 0.8 and 0.856 on the training data for which they were developed. AUC ranged from 0.766 to 0.883 on external and additional validations, using the baseline and intermediate data of the various trials.

Since TJC and SJC are clinically important measures for the rheumatologist and models without them would be considered to lack face validity, we additionally considered alternative models in which a change in TJC or SJC was forcibly included. These alternative models were developed to have these change-in-joint-count measures statistically

significant and maximal, in the sense that further inclusion of any new variable would make the joint count measures not statistically significant. Inclusion of such measures did not appear to hinder the "testing" performance of models.

Additionally, it was apparent that the CRP is important in any outcome developed. Further, it was clear that some subjective measure of disease activity (patient or physician), arthritis pain, quality of life, or functional limitation is also required in a response to treatment outcome for PsA. However, these measures may all be highly correlated with one another and thus attempting to explain the same variability in the data.

Factor analyses of difference measures (with or without inclusion of baseline values of these measures) showed evidence for clustering (high factor loadings) of the PtGDA and MDGDA measures, with the patient arthritis pain score, the SF-36PCS, and to a lesser extent the HAQ score clustered into 1 latent factor item. They also showed that HAQ and the SF-36PCS may cluster under a separate latent factor representing physical functioning, while TJC and SJC measures and the CRP measure also load highly on their own latent factor. In addition, a fifth latent factor was derived and indicated that the 2 SF-36 variables clustered together. These 5 latent factors explained about 64% of the variation in this data on difference measures (Table 2).

The results from the factor analyses and general observa-

Table 1. Summaries for 24-week changes from baseline for the various measures. (D\_ denotes differences between current and baseline values).

Training Difference Data (change from baseline)	Mean (SE)		t-value Drug vs Placebo	Unadjusted Z-value for OR		Effect Size		SRM		GES Drug vs Placebo	GUES
	Placebo	Drug		Drug vs Placebo	Drug vs Placebo	Placebo	Drug	Placebo	Drug		
D_TJC68	-4.066 (0.824)	-10.693 (0.835)	-5.64	0.955	-5.19	-0.2594	-0.7661	-0.3390	-0.8430	-0.5362	-0.8914
D_SJC66	-2.854 (0.572)	-6.368 (0.548)	-4.44	0.946	-4.17	-0.2800	-0.6793	-0.3429	-0.7644	-0.4220	-0.7651
D_HAQ	-0.118 (0.029)	-0.477 (0.034)	-7.95	0.183	-6.88	-0.2178	-0.7462	-0.2859	-0.9162	-0.7583	-1.1519
D_CRP	-0.043 (0.104)	-1.281 (0.158)	-6.44	0.611	-5.79	-0.0662	-0.5263	-0.0286	-0.5370	-0.6155	-0.8523
D_logCRP	-0.075 (0.058)	-1.393 (0.087)	-12.44	0.315	-9.15	-0.0992	-1.0721	-0.0893	-1.0652	-1.1894	-1.6600
D_MDGDA (Likert)	-0.333 (0.065)	-1.329 (0.069)	-10.37	0.360	-8.37	-0.5023	-1.8403	-0.3546	-1.2689	-0.9983	-1.4138
D_PtGDA (Likert)	-0.081 (0.068)	-1.069 (0.077)	-9.55	0.426	-7.99	-0.1123	-1.1643	-0.0810	-0.9168	-0.9090	-1.0753
D_Patient assessment of pain (Likert)	-0.109 (0.071)	-1.165 (0.081)	-9.72	0.435	-8.07	-0.1708	-1.3199	-0.1059	-0.9443	-0.9255	-1.1313
D_SF-36PCS*	1.544 (0.621)	9.636 (0.672)	8.82	1.096	7.47	0.1573	0.9688	0.1712	0.9618	0.8478	1.0685
D_SF-36 MCS*	0.317 (0.763)	2.492 (0.656)	2.31	1.023	2.29	0.0983	0.2220	0.0324	0.2551	0.2225	0.2548

\* Improvement over time is indicated by a positive increase in the score, as compared to the other measures where a decrease indicated improvement. SRM: Standardized response mean; GES: group effect size; GUES: Guyatt's effect size; TJC: tender joint counts; SJC: swollen joint counts; HAQ: Health Assessment Questionnaire; CRP: C-reactive protein; MDGDA: physician global assessment of disease activity; P+GDA: patient global assessment of disease activity; SF-36 PCS: Medical Outcomes Study Short-Form physical component; SF-36 MCS: SF-36 mental component.

Table 2. Factor analysis results for the differences in terms of the factor loadings.

	Factor 1	Factor 2	Loadings Factor 3	Factor 4	Factor 5
D_SJC66	0.139	0	<b>0.733</b>	0	0
D_TJC68	0.209	0.228	<b>0.632</b>	0.182	0
D_CRP	0.183	0.115	0	<b>0.520</b>	0.101
D_MDGDA	<b>0.490</b>	0.129	0.249	0.367	0
D_PtGDA	<b>0.860</b>	0.195	0.220	0.150	0.169
D_patient assessment of pain	<b>0.744</b>	0.233	0.176	0.262	0.127
D_HAQ	0.313	<b>0.868</b>	0.225	0.214	0.219
D_SF-36PCS	<b>-0.586</b>	<b>-0.528</b>	-0.225	-0.220	<b>0.523</b>
D_SF-36MCS	-0.159	-0.133	-0.136	-0.153	<b>-0.407</b>
Percentage of the variance explained by latent factors	23.3	13.6	13.3	7.2	6.1
Cumulative percentage	23.3	36.9	50.2	57.4	63.6

For abbreviations see Table 1. Bold type indicates significant loading.

tion therefore suggested the following 4 domains: (1) a subjective assessment (patient or physician) domain comprising disease, pain, and physical functioning representing disease severity; (2) a joint count domain; (3) an inflammatory marker domain (CRP); and (4) a general physical and psychological well-being domain. To capture the first domain, we derived a new variable that was the maximum of the difference measures for the MDGDA, PtGDA, patient assessment of arthritis pain, and for the HAQ score after rescaling by 5/3. This new variable was constructed in order to avoid including all these possibly highly correlated variables as main effects in analyses. We identified the second domain by the use of both the baseline and difference measures for TJC. We use the TJC to represent this domain rather than the SJC because TJC have been shown to be more reliably measured<sup>16,17</sup>. Further, models that included SJC measures did not provide any further discriminatory power over those excluding these measures. The fourth domain was identified by using both SF-36PCS and SF-36MCS. On fitting a logistic regression model to these 4 domains, we found that the SF-36 variables were not necessary (i.e., nonsignificant) and therefore were dropped on further refinement of the model. In the final model, we found that the joint count domain could be represented by only the “current” 24-week TJC since the baseline count did not contribute to discrimination between the treatment groups. The results are shown in Table 3. This model produced AUC and SE of 0.846 (0.019), 0.821 (0.035), 0.892 (0.024), and 0.826 (0.025) for the training dataset, the joint baseline and intermediate dataset from the other trial, and the joint baseline and intermediate dataset from the 2 trials represented in the training set, respectively.

*Multivariate analysis for percentage change measures.* When a multivariate logistic regression model, with automatic stepwise selection, was fitted to the percentage change measures, the model in Table 4 resulted. The AUC (with SE) were 0.831 (0.020) on the data the model was built on, and

Table 3. Logistic regression results based on domains.

	Estimate	Domain Model SE	z-value	p
Intercept	0.6935	0.1898	3.65	0.0003
C_TJC68	-0.0212	0.0083	-2.55	0.0108
D_CRP	-0.8936	0.1791	-4.99	< 0.0001
BL_CRP	-0.6061	0.1452	-4.17	< 0.0001
MAX.SUBJ.ASSESS	-0.9679	0.1625	-5.96	< 0.0001

C\_ denotes current; D\_ denotes difference; BL\_ denotes baseline; MAX.SUBJ.ASSESS: max (D\_MDGDA, D\_PtGDA, D\_patient assessment of pain, 5/3.D\_HAQ). For abbreviations see Table 1.

Table 4. Logistic regression results for percentage change measures.

	Estimate	Percentage Change Model SE	z-value	p
Intercept	-0.8958	0.1629	-5.50	< 0.0001
PC_TJC68	-0.5290	0.2313	-2.29	0.0222
PC_CRP	-0.3298	0.1430	-2.31	0.0211
PC_MDGDA	-1.6310	0.4155	-3.93	< 0.0001
PC_Patient assessment of change	-0.7936	0.3577	-2.22	0.0265
PC_HAQ	-0.6349	0.2787	-2.28	0.0227

PC\_ denotes percentage change. For abbreviations see Table 1.

0.836 (0.034), 0.851 (0.028), and 0.820 (0.026) for the joint baseline and intermediate data from all 3 trials for validation.

*Tree-based analyses.* Tree-based models offered no improvement to the logistic analyses consistent with the observation that no noticeable interactions were found among the various variables in the logistic models presented, and the tree-based models are specifically advantageous when interactions are present.

*The use of PASI.* As PASI was measured only on those patients who had more than 3% body surface area affected

by psoriasis and was not provided in 1 trial, it was not considered in the development and comparison of the statistical models. However, a preliminary exploration of the effects of including PASI measures to the index developed from the domain model of Table 3 revealed a statistically significant negative effect of an interaction of the index with the availability of baseline PASI on whether randomized to the active drug group in the training dataset. Thus the index developed from the Domain model was substantially reduced in its ability to discriminate between active drug and placebo when PASI is included. This was confirmed in the testing sets. In addition, the effects of the PASI measures on the randomized group outcome were highly statistically significant. This thus suggests that anti-TNF inhibitors have a major effect on psoriasis, which may support the argument for treating the skin aspect of the disease separately from other aspects when examining response to treatment.

*Dactylitis, enthesitis, and duration of morning stiffness.* Assessing the possible effect of including the dactylitis, enthesitis, and morning stiffness measures, had they been available for all the trials rather than for some, revealed that the effects of the “optimal” linear predictor index from Table 3 did not differ significantly (quantitatively) for those patients with or without these measures available. Further, for the data from those who have these measures recorded, the measures were not found to further discriminate, beyond the index alone, between those randomized to receive drug and those to placebo. As PASI and information on other elements of PsA were incomplete or absent, our models mainly reflect the joint aspects of the disease.

Finally, in order to define a binary response, for example, to treatment in a clinical trial, an investigation of the choice of thresholds for the indices (i.e., linear predictors) from either of the domain or percentage change models is needed. However, that was not investigated here.

## DISCUSSION

Despite the differences between PsA and RA, most drug trials have used as a primary outcome the ACR20 response, developed for RA and not validated for PsA. While the PsARC was developed specifically for PsA, it had not been validated prior to its first use in the sulfasalazine trial, and in that trial, did not function very well. Fransen, *et al*<sup>12</sup> demonstrated that the ACR20, PsARC, and the DAS all discriminated well between patients treated with etanercept and infliximab, but did not try to derive a new measure based on the data.

We aimed to derive a statistical model that would be more directly informative concerning response measures for PsA. We used the data collected in the recent phase III trials with anti-TNF agents, which have all shown improvement in signs and symptoms of the disease. We used the randomization to define response groups and used 3 approaches, a

logistic regression, a tree analysis, and a factor analysis, to develop discriminatory models. Moreover, we used data from 2 trials as the training set, and tested the results on the data from the third trial, as well as the interim results from the 2 studies used for the training set. Two models were developed based on PtGDA and MDGDA, patient reported outcomes, joint counts, and CRP. The first is based on differences between last visit and baseline variables, and includes current TJC, baseline and difference in CRP value, and the largest difference in the global disease activity scores or patient-reported outcomes. The development of this “difference” model was based on additional clinical input. This was because neither of the joint count measures entered the model purely on statistical considerations. This may reflect the high correlations between some of the other difference measures and the joint count measures and the fact that some of the measures included may act as mediators since they may be further along in the “causal pathway.” Because this “difference” model was influenced by formal analyses and clinical considerations, independent validation would be valuable.

The second model is based on the percentage change in variables and includes TJC (68 joints), CRP, MDGDA, patient assessment of pain, and HAQ. Both models provided high AUC, consistently above 0.8 for both the training and testing sets.

Most of the items included in the models are included in the core domains recommended by OMERACT<sup>13</sup>. However, the OMERACT core set does not include CRP, but does include PASI. PASI scores are highly discriminatory for treatment with biologics and dominated measures linked to arthritis for this purpose, suggesting that the skin should be considered as a separate domain. Although assessment of CRP was not included among the OMERACT core domains, it was included among the domains that should be measured. Moreover, acute-phase reactants have important prognostic implications in PsA, in terms of both progression of joint damage and mortality<sup>17,18</sup>. It should also be noted that all these items are reliable and feasible because they are all measured in clinical trials, and indeed are easily measured even in clinical practice.

The 2 models that provide effective discrimination between patients randomized to an anti-TNF agent or placebo need to be compared with the measures currently used to assess response, particularly the ACR20 and the PsARC, to further inform the development of response measures for PsA trials. Further, these models need to be assessed, based on different considerations, to determine whether they are useful in evaluating clinical response to treatment (in terms of functional outcomes such as damage progression) in individual patients.

## ACKNOWLEDGMENTS

The authors thank Abbott Laboratories, Amgen Inc., and Centocor Ortho Biotech Inc. for providing the data from their respective trials.

## REFERENCES

1. Gladman DD. Epidemiology. Psoriatic arthritis. In: Gordon GB, Ruderman E, editors. Psoriasis and psoriatic arthritis: an integrated approach. Heidelberg, Germany: Springer-Verlag; 2005:57-65.
2. Gladman DD, Helliwell P, Mease PJ, Nash P, Ritchlin C, Taylor W. Assessment of patients with psoriatic arthritis. A review of currently available measures. *Arthritis Rheum* 2004;50:24-35.
3. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
4. van der Heijde DM, van 't Hof M, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* 1993;20:579-81.
5. Clegg DO, Reda DJ, Mejias E, Cannon GW, Weisman MH, Taylor T, et al. Comparison of sulfasalazine and placebo in the treatment of psoriatic arthritis. A Department of Veterans Affairs Cooperative Study. *Arthritis Rheum* 1996;39:2013-20.
6. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet* 2000;356:385-90.
7. Kaltwasser P, Nash P, Gladman D, Mease P, Wolfgang J. Efficacy and safety of leflunomide in the treatment of psoriatic arthritis: results from the TOPAS Study. *Arthritis Rheum* 2002;46:3416-7.
8. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, et al. Etanercept treatment of psoriatic arthritis: safety, efficacy, and effect on disease progression. *Arthritis Rheum* 2004;50:2264-72.
9. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EH, et al, for the ADEPT Study Group. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: results of a double-blind, randomized, placebo-controlled trial. *Arthritis Rheum* 2005;52:3279-89.
10. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, et al. Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: results from the Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT). *Arthritis Rheum* 2005;52:1227-36.
11. Antoni C, Krueger GG, de Vlam K, Birbara C, Beutler A, Guzzo C, et al. Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. *Ann Rheum Dis* 2005;64:1150-7.
12. Fransen J, Antoni C, Mease PJ, Uter W, Kavanaugh A, Kalden JR, et al. Performance of response criteria for assessing peripheral arthritis in patients with psoriatic arthritis: Analysis of data from randomized, controlled trials of two TNF inhibitors. *Ann Rheum Dis* 2006;65:1373-8.
13. Gladman DD, Mease PJ, Strand V, Healy P, Helliwell PS, FitzGerald O, et al. Consensus on a core set of domains for psoriatic arthritis. OMERACT 8 PsA Module Report. *J Rheumatol* 2007;34:1167-70.
14. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;40:171-8.
15. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Proceedings of the Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 1973:267-81.
16. Gladman DD, Cook RJ, Schentag C, Feletar M, Inman RI, Hitchon C, et al. The clinical assessment of patients with psoriatic arthritis: results of a validation study of the SpondyloArthritis Research Consortium of Canada (SPARCC). *J Rheumatol* 2004;31:1126-31.
17. Gladman DD, Inman R, Cook R, Maksymowych W, Braun J, Davis JC, et al. International spondyloarthritis inter-observer reliability exercise — The Inspire Study: II. Assessment of peripheral joints, enthesitis and dactylitis. *J Rheumatol* 2007;34:1740-5.
18. Gladman DD, Cook RJ, Schentag C, Feletar M, Inman RI, Hitchon C, et al. The clinical assessment of patients with psoriatic arthritis: results of a validation study of the SpondyloArthritis Research Consortium of Canada (SPARCC). *J Rheumatol* 2004;31:1126-31.