Informing Response Criteria for Psoriatic Arthritis (PsA). II: Further Considerations and a Proposal — The PsA Joint Activity Index

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

ABSTRACT. ***Objective.*** To develop a recommended measure of response for use in psoriatic arthritis (PsA) clinical trials and observational cohort studies reflecting joint involvement.

Methods. Previously, we used data from phase III randomized placebo-controlled trials of anti-tumor necrosis factor (TNF) agents to determine models, based primarily on statistical considerations but with some clinical input when necessary, that best distinguish drug-treated from placebo-treated patients. For the same data, we examine response criteria currently used for PsA and logistic regression models based on the individual components of these response criteria. Comparison with our previously developed models, based primarily on statistical consideration, is made.

Results. A simplified score, the PsA Joint Activity Index (PsAJAI), based on components of the ACR30, performed better than the ACR20 and PsARC, and comparable to our previously developed models. The PsAJAI is a weighted sum of 30% improvement in core measures with weights of 2 given to the joint count measure, the C-reactive protein laboratory measure, and the physician global assessment of disease activity measure. Weights of 1 should be given to the remaining 30% improvement measures including pain, patient global assessment of disease activity, and the Health Assessment Questionnaire.

Conclusion. We recommend the PsAJAI be used as an outcome measure for assessing joint disease response in PsA clinical trials. (J Rheumatol First Release Oct 15 2010; doi:10.3899/jrheum.100479)

Instruments used in clinical trials in psoriatic arthritis (PsA) to date include the disease activity score (DAS) and DAS28 and the American College of Rheumatology (ACR) 20% response criteria (ACR20) developed for rheumatoid arthritis (RA)[1,2,3]. The Psoriatic Arthritis Response Criteria (PsARC), a composite instrument originally developed for a sulfasalazine study in PsA, has also been used[4]. None have been validated for PsA prior to their use in clinical trials,

where both the ACR20 and the PsARC demonstrated efficacy of anti-tumor necrosis factor (TNF) agents, as well as leflunomide, in patients with PsA[5,6,7,8,9,10]. Indeed, in a previous investigation to compare the responsiveness and discriminative capacity of various response criteria, the Disease Activity Score (DAS) and core-set measures in PsA patients with peripheral arthritis from two phase II randomized placebo-controlled trials of TNF inhibitors, it was found that the ACR20 performed better than the PsARC in discriminating active drug from placebo, but both were found to be useful response measures for PsA[11].

Previously, we used data from phase III randomized placebo-controlled trials of anti-TNF agents to determine models, based primarily on statistical considerations but with some clinical input when necessary, that best distinguish drug-treated from placebo-treated patients[12]. We used as a training set the data from baseline and 24 weeks of 2 anti-TNF trials, and then tested the results on the baseline and interim data of the third trial (external validation), as well as the baseline and interim data for the first 2 trials (additional validation)[12]. Two models were derived: a domain model 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 (PAIN), and the overall Health Assessment Questionnaire (HAQ) score. The second model was based on percentage change from baseline and included TJC68, CRP, physician GDA, PAIN, and HAQ. Both models provided high area under the curve (AUC) for receiver-operating characteristic curves of at least 0.8 for both the training and testing sets. In this investigation, the percentage change model had comparable properties to the domain model, which was based on differences.

In spite of statistical concerns with percentage change measures, their use is deeply entrenched in currently used instruments of response. Since our aim was to examine the performance of these current instruments in light of the results from our earlier report[12], we have developed a specific instrument based on results that use percentage change information.

MATERIALS AND METHODS

Data. The data came from 3 recent clinical trials of TNF inhibitors in patients with PsA[7,8,10]. In total, 366 patients were randomized to the placebo arms of the trials and 354 to the drug arms. We extracted a common combined dataset from the 3 trials' data that could be used to investigate measures of responsiveness for PsA based on information available in all 3 trials[12]. For tender and swollen joint counts we chose variables in the combined dataset that recorded the 68 tender joint counts (TJC68) and the 66 swollen joint counts (SJC66), which had to be derived from 78/76 joint counts for one trial. Likert-type variables were derived for patient and physician global assessment of disease activity (PtGDA and MDGDA, respectively).

Data from baseline and 24 weeks of 2 trials were used as the training set (i.e., the set of data from which models are built), whereas baseline and interim (12- or 14-week) data from the 3 trials were used as testing sets (i.e., sets of data from which the models built are validated).

Finally, known improvement/response criteria (yes/no) indicator variables were constructed. These were the ACR improvement criteria with 20%, 30%, and 40% cutoff points (denoted ACR20, ACR30, ACR40) and the PsARC. The additional levels of response to the ACR criteria (i.e., 30% and 40%) were introduced, as in PsA a placebo response could possibly be as high as 30%. The EULAR definition for responsiveness in RA based on the DAS28 is not considered here due to the lack of individual joint-level data in one of the trials. However, our domain model[12], based on differences and developed specifically for PsA, is of the same nature as the DAS-based response instrument for RA.

The definition of response for the ACR20 instrument is at least a 20% improvement in tender and swollen joint counts and at least a 20% improvement in 3 of the remaining 5 core measures: CRP, MDGDA, PtGDA, PAIN, and HAQ. The ACR30 and ACR40 are defined similarly to the ACR20, but with the 20% core measures replaced by the corresponding 30% and 40% core measures, respectively. Measured response under PsARC is defined as an improvement in at least 2 of the 4 core measures (TJC, SJC, MDGDA, and PtGDA), one of which has to be either tender joint count or swollen joint count, and with no worsening in any of the 4. Improvement in the 4 core measures of PsARC is defined as 30% improvement in TJC and SJC, and a decrease by one category on the Likert scales for physician and patient global assessments of articular disease (i.e., disease activity).

Statistical methods. Although not a "gold standard" for responsiveness, the treatment indicator of whether randomized to the placebo or drug arms of

a trial was used as a proxy measure for nonresponse or response. These biologic therapies have been shown to be dramatically more effective in treating the symptoms of PsA than earlier disease-modifying antirheumatic drug therapies.

Based on logistic regression (LR) models including all of the yes/no variables used in constructing the ACR improvement criteria based on 20%, 30%, and 40% improvement, we developed corresponding responsiveness indices (i.e., LR-ACR20, LR-ACR30, and LR-ACR40, respectively) for predicting whether or not a patient was randomized to receive drug in the training dataset. A responsiveness index (LR-PsARC) was similarly developed based on the yes/no variables used in the construction of the PsARC. These responsiveness indices are derived from the linear predictors of the logistic regression models.

These linear predictors, and those of the domain and percentage change models[12], were used to form overall binary decision response indicators that defined whether or not a patient responded. The cutoff points, c1 and c2, used for the dichotomization were chosen to set the specificity of the linear predictor equal to (or approximately equal to) the specificity of its corresponding ACR or PsARC criteria, and to be equal to the mean of the linear predictor, respectively. These binary indicators, defined in a consistent way for all investigations, illustrate how potential yes/no response indicators can be formed from their linear predictor scores.

Evaluation of the various response measures defined (new and existing) was based, where appropriate, on sensitivity and specificity, deviances and degrees of freedom, z values, and area under the receiver-operating characteristic (ROC) curve.

RESULTS

Investigation of currently used instruments for responsiveness. Table 1 presents sensitivity and specificity results for the ACR improvement criteria with 20%, 30%, and 40% cutoff points and the PsARC. In addition, it presents the results of univariate logistic regression models for discriminating between treatment groups using these yes/no response indicators.

The sensitivity of the PsARC was 76%, which was the highest among the 4 established response criteria. However, the specificity of the PsARC was 67%, the lowest among the 4 response criteria. Among the ACR criteria, as expected, the ACR20 had the highest sensitivity, while the ACR40 had the highest specificity. All 4 criteria were highly significant in discriminating active drug from placebo (see z values in Table 1).

Responsiveness indices obtained through reexamining the individual components of currently used instruments. The results for the logistic regression models with all the individual yes/no components used in constructing the PsARC and the ACR20%, ACR30%, and ACR40% criteria as explanatory variables are shown in Table 2. All variables in the LR-PsARC index were found to be statistically significant at the 5% level. Improvements in these variables all increased the probability of having been randomized to the active-drug group. The area under the ROC curve for the LR-PsARC index was 0.78. When this index was dichotomized at the cutpoints c1 and c2, taking values 0.289 and 0.033, respectively, which were the chosen thresholds for indicating a positive response, the sensitivity and specificity were very similar to that obtained from PsARC (i.e., sen-

Response Criterion	Drug (n = 219), Sensitivity	Placebo (n = 202), Specificity	Logistic Regression (LR) Models				
			Estimate	SE	z	Residual Deviance	Residual Degrees of Freedom
PsARC	0.76	0.67	1.89	0.22	8.65	499.5	419
ACR20	0.58	0.82	1.85	0.23	8.07	508.3	419
ACR30	0.50	0.89	2.09	0.26	7.95	503.5	419
ACR40	0.37	0.95	2.42	0.35	6.86	511.6	419

ACR: American College of Rheumatology; PsARC: Psoriatic Arthritis Response Criteria.

sitivity between 0.76 and 0.81, and specificity between 0.66 and 0.68), with both LR-PsARCc1 and LR-PsARCc2 doing marginally better overall (based on the summation of the sensitivity and specificity) than PsARC. On assessing the residual deviances and corresponding residual degrees of freedom of the LR-PsARC index and binary counterparts with the original PsARC instrument we observed that the discriminating abilities of the former surpassed those of the latter.

For the 3 LR-ACR indices (20 to 40), 3 of the 6 core measures included in their construction were found always to be statistically significant. These were the percentage improvement indicators for both tender and swollen joints, CRP, and the physician global assessment of disease activity, with each indicating that at least 20%, 30%, or 40% improvement increases the likelihood of having been randomized to the active-drug group. The HAQ measure was statistically significant in 2 of the 3 ACR logistic regression models (LR-ACR20 and LR-ACR40). Patient measures of pain and global disease activity generally contributed little to the model fit. Surprisingly, although not statistically significant, the effect estimate for the 40% improvement measure for patient global assessment of disease activity was, counterintuitively, negative. This was not the case for the LR-ACR20 and LR-ACR30 indices.

The LR-ACRc1 and LR-ACRc2 yes/no response indicators had sensitivities significantly higher than those of the original ACR criteria at 20%, 30%, and 40% improvement given in Table 1. However, the specificities of the LR-ACRc2 measures were lower than the original ACR measures. Nevertheless, the overall best performing dichotomizations (in terms of largest values for the summation of the sensitivity and specificity) were from the LR-ACR30c1 and LR-ACR30c2 binary indicators.

The areas under the ROC curves for the 4 LR indices ranged from 0.78 to 0.86 on the training data (Table 2), with the largest 2 areas under the ROC curves coming from the LR-ACR30 and LR-ACR40 indices. These 2 indices also had the smallest residual deviances among the 3 LR-ACR indices. The external and additional validation results for these 4 new responsiveness indices (LR-PsARC, LR-ACR20, LR-ACR30, and LR-ACR40) are presented in Table 3. All indices appear to be robust, in particular the

LR-ACR indices. Overall, the LR-ACR30 index performed best among the 4. In addition, the LR-ACR30 dichotomizations (i.e., LR-ACR30c1 and LR-ACR30c2) had more significant z values (Table 2) than the ACR20, ACR30, and ACR40, the LR-ACR20 dichotomizations, and the LR-ACR40c1 dichotomization. Thus, there may be more discriminatory power in the 30% improvement measures that comprise the ACR30 criteria than in the 20% improvement measures that comprise the ACR20 criteria. Further, our analyses suggest that a more optimal way of constructing a response instrument for PsA can be derived than through the logical (or Boolean or tree-based) definitions of the PsARC or ACR criteria, although simplicity of use should also be considered when constructing such an instrument.

Proposal for a simplified PsA joint activity index (PsAJAI). For further investigation of whether a "better" response index could be derived, which would be simple to apply and perform well in a randomized controlled trial or clinical setting, we examined the ACR30 instrument and the LR-ACR30 index further. We found when considering the ACR30 instrument that no significant improvement on the ACR30 could be made through simply altering the original definition of improvement in response to some other logical (Boolean or tree-based) combination of the seven 30% improvement measurements. This confirms that interactions among core measure variables are not important for deriving a response instrument for PsA, as seen in our earlier publication[12].

However, on assessment of the estimates obtained from the percentage improvement measures in the LR-ACR30 model (or equivalently, the coefficients of the LR-ACR30 index), and with ease of clinical usage and clinical assessment of importance in mind, we were able to adapt the relative weighting of these measures to obtain a simplified LR-ACR30 index, denoted the PsAJAI. We defined this new simplified score as follows:

PsAJAI = 2 × 30% ↓ JNT + 2 × 30% ↓ CRP + 2 × 30% ↓ MDGDA + 30% ↓ PTGDA + 30% ↓ PAIN + 30% ↓ HAQ

which results from calculating the 30% improvement indi-

Table 2. ACR and PsARC results using the derived models.

Response Index	Cutoff Point	Sensitivity	Specificity	Logistic Regression (LR) Models					
				Estimate	SE	z	Residual Deviance	Residual Degrees of Freedom	Area Under ROC Curve
LR-PsARC index*							480.5	417	0.78
Intercept				−1.27	0.24	−5.20			
PC30 JNT IMPR				0.87	0.28	3.08			
MDGDA-1 cat				−0.85	0.20	−4.23			
PtGDA-1 cat				−0.69	0.17	−4.04			
LR-PsARCc1	0.289	0.79	0.68	2.07	0.22	9.24	485.5	419	
LR-PsARCc2	0.033	0.81	0.66	2.12	0.23	9.31	482.6	419	
LR-ACR20 index**							442.1	414	0.82
Intercept				−2.39	0.28	−8.55			
PC201 JNT				0.69	0.25	2.76			
PC201 CRP				1.24	0.25	5.06			
PC201 MDGDA				1.09	0.26	4.13			
PC201 PTGDA				0.19	0.30	0.64			
PC201 PTASSPAIN				0.48	0.30	1.60			
PC201 HAQ				0.58	0.25	2.27			
LR-ACR20c1	0.597	0.68	0.82	2.28	0.23	9.76	469.3	419	
LR-ACR20c2	0.076	0.74	0.73	2.05	0.22	9.27	485.3	419	
LR-ACR30 index***							414.4	414	0.84
Intercept				−2.28	0.26	−8.84			
PC301 JNT				0.82	0.27	3.07			
PC301 CRP				1.59	0.25	6.34			
PC301MDGDA				1.00	0.28	3.62			
PC301PtGDA				0.58	0.37	1.57			
PC301 PtASSPAIN				0.25	0.36	0.68			
PC301 HAQ				0.51	0.27	1.92			
LR-ACR30c1	0.872	0.63	0.89	2.65	0.27	9.98	449.4	419	
LR-ACR30c2	0.135	0.73	0.84	2.63	0.24	10.79	437.3	419	
LR-ACR40 index†							393.7	414	0.86
Intercept				−2.08	0.24	−8.78			
PC401 JNT				0.79	0.28	2.86			
PC401 CRP				1.95	0.26	7.63			
PC401 MDGDA				1.10	0.30	3.62			
PC401 PtGDA				−0.57	0.48	−1.19			
PC401 PTASSPAIN				1.50	0.46	3.24			
PC401 HAQ				0.82	0.28	2.87			
LR-ACR40c1	1.185	0.54	0.95	3.13	0.35	8.90	445.7	419	
LR-ACR40c2	0.224	0.72	0.83	2.53	0.24	10.50	447.0	419	

* PC30 JNT SWELL CNT — Joint swelling count: improvement = decrease by at least 30%, worsening = increase by at least 30%, otherwise same. PC30 JNT PAIN CNT — Joint pain/tenderness count: improvement = decrease by at least 30%, worsening = increase by at least 30%, otherwise same. PC30 JNT IMPR = (Improvement in PC30 JNT SWELL CNT) or (Improvement in PC30 JNT PAIN CNT). MDGDA — 1 cat — Physician global assessment of articular disease (disease activity): improvement = decrease by at least one category (−1), worsening = increase by at least one category (1), same = no category change (0). PTGDA-1 cat— Patient global assessment of articular disease (disease activity): improvement = decrease by at least one category (−1), worsening = increase by at least one category (1), same = no category change (0).** PC201 JNT SWELL, PC201 JNT PAIN, PC201 CRP, PC201 MDGDA, PC201 PTGDA, PC201 PTASSPAIN, and PC201 HAQ are indicators for at least 20% improvement in swollen joint count, tender/painful joint count. C-reactive protein (CRP) measurement, physician global assessment of disease activity, patient global assessment of disease activity, patient assessment of pain and Health Assessment Questionnaire (HAQ) score. PC201 JNT = (PC201 JNT SWELL = 1) and (PC201 JNT PAIN = 1). *** PC301 JNT SWELL, PC301 JNT PAIN, PC301 CRP, PC301 MDGDA, PC301 PTGDA, PC301 PTASSPAIN, and PC301 HAQ are indicators for at least 30% improvement in swollen joint count, tender/painful joint count, C-reactive protein measurement. Physician global assessment of disease activity. Patient global assessment of disease activity. Patient assessment of pain and HAQ score. PC301 JNT = (PC301 JNT SWELL = 1) and (PC301 JNT PAIN =1). † PC401 JNT SWELL, PC401 JNT PAIN, PC401 CRP, PC401 MDGDA, PC401 PTGDA, PC401 PTASSPAIN, and PC401 HAQ are indicators for at least 40% improvement in swollen joint count, tender/painful joint count, CRP measurement, Physician global assessment of disease activity. Patient global assessment of disease activity, patient assessment of pain and HAQ score. PC401 JNT = (PC401 JNT SWELL = 1) and (PC401 JNT PAIN = 1).

cators of the 7 core measures of the ACR, giving a higher weight of 2 to the joint, the laboratory measure, and the physician global assessment indicators, and a lower weight of 1 to the remaining patient assessment indicators, and then summing to get a score out of 9. The discriminatory power of the PsAJAI is assessed relative to the LR-ACR30 index.

We found that the area under the curve for the PsAJAI was 0.83 compared to 0.84 obtained from the LR-ACR30. Additionally, if we choose a cutpoint ≥ 5 to decide whether a patient is to be predicted as belonging to the active-drug

Table 3. Comparison of existing response criteria in psoriatic arthritis.

| | Area Under the Receiver-operating Characteristic Curves (SE) | | |
Response Index	Trial 1	Trial 2	Trial 3
LR-PsARC	0.768 (0.040)	0.763 (0.034)	0.784 (0.028)
LR-ACR20	0.800 (0.037)	0.848 (0.028)	0.837 (0.025)
LR-ACR30	0.801 (0.037)	0.864 (0.027)	0.853 (0.024)
LR-ACR40	0.763 (0.040)	0.882 (0.025)	0.851 (0.024)

LR: logistic regression.

group, then the sensitivity and specificity of this dichotomized version of PsAJAI are 0.74 and 0.84, respectively, which when summed is greater than the corresponding sums for the already established instruments (Table 1). This would thus indicate an increase in discriminatory capacity of the PsAJAI over these currently used responsiveness criteria.

Assessment of PsAJAI against ACR instruments. Cross-tabulations of the ACR30 with our PsAJAI dichotomized at a cutpoint of 5 (as above) are shown in Table 4 for the training data in each of the treatment arms separately. It is apparent that more patients are deemed responders with the PsAJAI (using cutpoint of 5), and that the majority are in the active-drug group. Hence, as suggested, it provides a possible advantage over the ACR30 improvement criteria.

Table 5 examines the reason why the 65 participants who achieved a PsAJAI response did not also achieve an ACR30 response. It breaks down this group of 65 responders by stratifying first on whether or not a 30% improvement in the joint count component was seen, and then by either the combination or number of other core measures that showed a 30% improvement. Of these 65 patients (irrespective of treatment group), 42 (64.6%) did not have a 30% improvement on the joint count component and 23 (35.4%) did have a 30% improvement on this component.

For the 23 who had a 30% improvement on the joint count component and therefore also had improvements in exactly 2 of the other 5 core measures, 12 (52.2%) had improvement on the CRP component, 18 (78.3%) had improvement on MDGDA, 2 (8.7%) on PtGDA, 3 (13.0%) on PAIN, and 11 (47.8%) on HAQ.

For the 42 patients who did not have a 30% improvement on the joint count component but were responders on our PsAJAI, 15 satisfied all the remaining 5 core measures making up the ACR30 criteria; 20 satisfied 4 of the remaining 5; and the last 7 satisfied 3 of the remaining 5, 2 of which were always CRP and MDGDA. Besides always satisfying the CRP and MDGDA improvement measures, 6 of these 7 patients also satisfied the HAQ improvement measure. The remaining individual satisfied the PAIN improvement measure instead of the HAQ improvement measure. Further, there was only one individual in this group of 42 patients who did not respond positively on the MDGDA improvement measure, and only 4 out of the 42 patients did not respond positively to the CRP improvement measure.

Further, of these 42 patients, only 9 patients had a 20% improvement in swollen joint count, and only 8 patients had a 20% improvement in tender joint count by 24 weeks. The median percentage improvement in tender joint counts for these 42 individuals was –29.7% (interquartile range –75%, 0%). The median percentage improvement in swollen joint counts for these 42 patients was –12.7% (IQR –44.6%, 12.6%).

The cross-tabulation results of Table 4 and the closer

Table 4. Comparison between ACR criteria and Psoriatic Arthritis Joint Activity Index (PsAJAI).

| | | PsAJAI ≥ 5 | |
		No	Yes
Placebo Group			
ACR30 criteria	No response	169	11
	Response	0*	22
Drug Group			
ACR30 criteria	No response	56	54
	Response	0*	109

* The reason for the zero count corresponding to responding on ACR30 but with an answer of "no" on the PsAJAI indicator is that when ACR30 declares a patient a responder this implies that PsAJAI has to be ≥ 5, because of the way the ACR30 response is constructed.

Table 5. Responders on PsAJAI but not on ACR30.

	Frequency (%)
Group with 30% improvement on joint count	
Exactly 2 other core measures showing 30% improvement	
MDGDA and HAQ	8 (34.8)
MDGDA and CRP	7 (30.4)
MDGDA and PAIN	2 (8.7)
MDGDA and PtGDA	1 (4.4)
CRP and HAQ	3 (13.0)
CRP and PAIN	1 (4.4)
CRP and PtGDA	1 (4.4)
Total	23 (100)
Group without 30% improvement on joint count	
No. of the remaining 5 core measures showing 30% improvement	
5 out of 5	20 (47.6)
4 out of 5	15 (35.7)
3 out of 5	7 (16.7)
Total	42 (100)

Table 6. Comparison of models.

Response Indicator	Cutoff Point	Sensitivity	Specificity	Logistic Regression (LR) Models				
				Estimate	SE	z	Residual Deviance	Residual Degrees of Freedom
Domain model c1	0.600	0.63	0.89	2.63	0.27	9.92	451.4	419
Domain model c2	−0.024	0.83	0.72	2.55	0.24	10.67	444.0	419
PC1 model c1	0.674	0.60	0.89	2.52	0.26	9.51	463.4	419
PC1 model c2	0.078	0.74	0.75	2.16	0.22	9.61	476.5	419

inspection above of the 65 patients suggest the reason that joint counts do not enter into logistic regression models developed through an automatic selection approach[12]. In addition, a cross-tabulation of the ACR20 improvement response with our PsAJAI dichotomized at a cutpoint of 5 showed that 39 out of 45 patients (86.7%) classified as not responding using the ACR20, but responding on PsAJAI, were in the active-drug group, while 9 of the 12 patients (75%) classified as responding on the ACR20 but not on our PsAJAI indicator were in the placebo group. This suggests a potential advantage of the PsAJAI over the ACR20 as well as the ACR30 criteria.

Domain and percentage change models. In our first report[12], we found that the areas under the curves for our domain and percentage change models developed based on primarily statistical considerations were 0.821, 0.892, and 0.826 (domain) and 0.836, 0.851, and 0.820 (percentage change) under external and additional validations using the baseline and interim data from the 3 trials. A comparison of the areas under the curves in Table 3 for our LR-ACR30 index with the above areas under the curves shows no strong evidence against using the LR-ACR30 index and subsequently the PsAJAI as a response index for clinical trials in PsA.

Additionally, we compared our previously derived domain and percentage change models to the various ACR and PsARC criteria and binary indicators. This comparison was performed on the same 421 patients used in Tables 1 and 2. Table 6 presents the results for these previously derived models based on dichotomization using the same strategy employed earlier for choosing cutpoints for the LR-ACR30 index. It is apparent that LR-ACR30c1 dichotomization produces sensitivity results similar to those for the domain and percentage change models dichotomized under the same "c1" strategy (Table 2). Additionally, both the LR-ACR30c2 dichotomization and the PsAJAI dichotomization at the previously chosen cutpoint of 5 produced specificities (both 0.84) that were better than those obtained from the "c2" dichotomization of the domain and percentage change models. They produced sensitivities lower than or equal to (both 0.74) those from the "c2" dichotomizations of the domain and percentage change models, respectively. These comparisons again suggest that the LR-ACR30 index and the simplified PsAJAI version of it perform comparably to those models, whether based on percentage change or dif-

ference, that were derived on the basis of primarily statistical considerations. However, the PsAJAI is easier to implement in clinical practice.

DISCUSSION

Trials of new therapies in PsA have generally incorporated outcome measures from other conditions, e.g., rheumatoid arthritis, psoriasis, and ankylosis spondylitis, with the exception of the PsARC. Indeed, the primary outcome measure in each of the major trials has been the ACR 20% response. The ACR20 was previously shown to function well in phase II trials in PsA[11].

We aimed to determine whether the response criteria that have been used in drug trials in PsA to date are indeed optimal. We have demonstrated in this exercise that the ACR20 performs well as a composite measure of disease response. However, detailed analysis suggests that the ACR30 outperforms the ACR20, and further that a better outcome measure of response may be the PsAJAI. Indeed, based on the principles of face validity, parsimony, and clinical simplicity, we would recommend that the PsAJAI be used to define response in PsA trials. This essentially involves a 6-variable checklist of 30% improvement indicators with a weight of 2 for joints, laboratory measurement, and physician components, and a weight of 1 for the remaining 3 patient components, global assessment of disease activity, pain, and HAQ. The PsAJAI was applied to the results of the ACCLAIM trial, where a response rate of 75.6% was noted, higher than the response of 70% obtained by PsARC and similar to the ACR20 response of 78%[13].

Based on our study we recommend this simple weighted sum of the 30% core measures, the PsAJAI, can be used as a joint response index and its properties can be further investigated in other datasets.

REFERENCES
1. 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.
2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EH, et al, for the ADEPT Study Group. Adalimumab in the Treatment of patients with moderately to severely active psoriatic arthritis: results of the ADEPT trial. Arthritis Rheum 2005; 52:3279-89.

9. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider E, 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.

10. 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.

11. 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.

12. Gladman D, Tom BDM, Mease PJ, Farewell VT. Informing response criteria for psoriatic arthritis. I: Discrimination models based on data from 3 anti-TNF randomized studies. J Rheumatol 2010;37:1892-7.

13. Gladman DD, Psaradellis F, Illouz O, Sampalis JS. Evaluation of response using the psoriatic arthritis joint activity index scoring tool in patients treated with adalimumab: Post hoc analysis of the ACCLAIM study [abstract]. J Rheumatol 2009;36:2571.