SYNOPSIS

SPONSOR:	**Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc.**	
COMPOUND NAME:	**MK-2155 (infliximab) or MK-8259 (golimumab)**	
INDICATION:	**Ankylosing Spondylitis**	
PROTOCOL TITLE:	A Prospective Observational Study to Evaluate the Relationship between Disease State and Change in Quality of Life in Ankylosing Spondylitis Patients treated with Remicade® (infliximab) or Simponi® (golimumab**)**	
TRIAL IDENTIFIERS:	Protocol Number:	194-00
TRIAL CENTERS:	This trial was conducted at 147 sites in 15 participating countries: 11 in Belgium, 5 in Bulgaria, 3 in Croatia, 5 in Czech Republic, 3 in Estonia, 21 in France, 32 in Germany, 6 in Greece, 11 in Hungary, 21 in Italy, 6 in Portugal, 7 in Romania, 6 in Russian Federation, 1 in Slovenia, and 9 in United Kingdom	
DESIGN:	This study was a multinational, prospective observational cohort study conducted in adult patients (>18 years of age) diagnosed with definite ankylosing spondilytis (AS) (according to the modified New York criteria) and newly prescribed with infliximab or golimumab in a regular clinical practice setting. The treatment with infliximab or golimumab started after patients met the eligibility criteria. Eligible patients received either infliximab or golimumab per the usual standard of care of the investigator during the follow-up period. Patients were individually followed for approximately six months with data collection at baseline (pre-treatment), three months and six months. Patient demographics, clinical characteristics, disease activity, health care resource utilization, and health-related quality of life (HRQoL) were summarized at these three timepoints. Baseline data (age, gender, race, medical history, and disease characteristics) were extracted from medical records available at each treatment center or collected at the time of enrollment through direct patient interview. No additional interventional tests or medical procedures (eg. blood samples, X-ray, or other technical investigations) were performed as a part of this study. If any data element was not available, it was reported as missing.	

<table>
<tr><td>DESIGN:
(Cont.)</td><td colspan="2">Efficacy parameters: Patient-reported outcomes (PROs) included Short Form 36 (SF-36) (to assess HRQoL); Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylolitis Functional Index (BASFI), Patient Global Assessment (PGA) of disease activity, PGA of Pain (Total back pain); and Work Productivity and Activity Impairment: Ankylosing Spondylitis (WPAI:SpA) (to collect data on work productivity and activity impairment). Additional parameters collected included data on health care utilization (use of concomitant medications, hospitalizations, and emergency room/outpatient visits), C-Reactive protein (CRP), and erythrocyte sedimentation rate (ESR).

Safety parameters: Sponteneously reported serious and non-serious adverse events (AEs) during six months of treatment period.

There were no comparisons made between infliximab and golimumab and the data for both agents were combined in the analyses. As such, no stratification was made according to the use of infliximab or golimumab.</td></tr>
<tr><td></td><td>Planned duration of main phase:
Planned duration of run-in phase:
Planned duration of extension phase:</td><td>6 months

not applicable

not applicable</td></tr>
</table>

Objectives	**Primary:** Develop an algorithm, using baseline (or pre-anti-tumor necrosis factor [TNF] treatment) parameters (demographic, clinical, AS severity) to predict change in HRQoL from baseline to six months, as measured by SF-36 Physical Component Summary (PCS) at six months, in AS patients newly treated with infliximab or golimumab. **Secondary:** 1. Develop an alternative algorithm using baseline parameters to predict change in HRQoL, as measured by the SF-36 Mental Component Summary (MCS) 2. Study the association between change in disease severity and change in HRQoL, as measured by SF-36 PCS, from baseline to three months and six months 3. Describe health care resource utilization and work productivity during treatment with infliximab or golimumab (at baseline, three months, and six months) 4. Compare health resource utilization and work productivity loss during treatment with infliximab or golimumab for SF-36 PCS responders and non-responders in terms of HRQoL, using data from baseline and six months 5. Investigate the external validity of a published clinical algorithm Vastesaeger et al.[1], developed from clinical trials, that identifies patients who respond to AS therapy, as measured by several clinical measures. The algorithm was validated using data at baseline and six months.	
Hypotheses	This is an observational study; no specific hypothesis is tested.	
Treatment groups	Infliximab	212 Subjects (22.0%)
	Golimumab	751 Subjects (78.0%)

Endpoints and definitions	Primary Efficacy Endpoint	Algorithm using baseline parameters that predicted change in HRQoL as measured by SF-36 PCS	An algorithm to predict change in HRQoL (measured by SF-36 PCS) from baseline to six months using baseline parameters that included demographic (age, gender), clinical (symptom duration, Human Leukocyte Antigen B27 [HLA-B27] genotyping, enthesitis score, CRP), AS disease severity (BASDAI score, BASFI score, Ankylosing Spondylitis Disease Activity Score [ASDAS] score, PGA of disease, PGA of pain), and other (number of comorbidities at baseline). Patients achieving an improvement in SF-36 PCS score of ≥ five points from baseline to six months were considered to be PCS responders; those with no improvement or < five points on the SF-36 PCS score were classified as PCS non-responders.
	Secondary Efficacy Endpoints	Alternative algorithm using baseline parameters that predicted change in HRQoL as measured by SF-36 MCS	An alternative algorithm to predict change in HRQoL (measured by SF-36 MCS) from baseline to six months using baseline parameters that included demographic (age, gender), clinical (symptom duration, HLA-B27 genotyping, enthesitis score, CRP), AS disease severity (BASDAI score, BASFI score, ASDAS score, PGA of disease and PGA of pain), and other (number of comorbidities at baseline). An improvement in SF-36 MCS score of ≥ five points from baseline to six months was considered as MCS responders; no improvement or < five points on the SF-36 MCS score was considered to be MCS non-responders.
		Association beween change in Disease Severity and change in HRQoL as measured by SF-36 PCS	The association between change in disease severity and change in HRQoL using SF-36 PCS response, from baseline to three months and six months. Change in BASDAI score, BASDAI50 response, ASDAS clinically important improvement, ASDAS major improvement, Assessment in SpondyloArthritis (ASAS)20, ASAS40 and ASAS partial remission were presented stratified by SF-36 PCS responders and PCS non-responders using the cutoff point of five at six months.

		Health Care Resource Utilization and Work Productivity during treatment with Infliximab and Golimumab	Descriptive statistics for health care resource utilization (acute care/ hospitalization /outpatient care, and concomitant medication usage for AS) and overall work productivity/activity impairment due to AS (measured by WPAI-SpA) were presented by overall, by SF-36 PCS responder/ non-responder at baseline, three months, and six months.
		External Validity of Clinical Algorithm to identify patients who responded to AS therapy	Investigate, in the QUO-VADIS study cohort, the validity of the clinical algorithm for identifying patients who will respond to AS therapy developed by Vastesaeger et al.[1] In this study, multivariate logistic regression analyses were performed for the outcome variables BASDAI50 response, ASDAS clinically important improvement, ASDAS major improvement, ASDAS inactive disease, ASAS20 response, and ASAS partial remission at six months using the following predictor variables described in the paper by Vastesaeger et al.[1], collected at baseline, to identify patients likely to respond to AS therapy: - Age: >40 vs. ≤40 - HLA-B27 genotype: Positive vs. negative - BASFI score: >6.5, ≤6.5 and >4.5, ≤4.5 - Berlin enthesitis score: =0 vs. >0 - CRP: <0.6, ≥0.6 and ≤2.0, >2.0
	Safety Analysis	Adverse Events	The incidences of spontaneously reported serious and non-serious AEs during the follow-up period
Database lock	13-JUL-2015	Trial status	21-NOV-2012 first subject first visit *to* 18-JUN-2015 last subject last visit. The study is completed.

RESULTS AND ANALYSIS	**General Principles:** Statistical analyses were conducted using SAS version 9.4 and Classification And Regression Tree (CART) model (Salford Systems). For continuous variables, the number of patients, arithmetic mean, standard deviation (SD), median, and minimum and maximum were presented. For categorical variables, the number of patients and percentage in each category were presented. The number and percentage of patients with missing values in every continuous and categorical variable were summarized. To calculate the percentage of patients with missing values, the denominator included the number of patients with missing values. The All Treated analysis set was defined as all subjects enrolled in the study who received at least one dose of study treatment. This group includes patients who discontinued the study or switched therapy during follow-up. Those who switched therapy were considered to be discontinuers. The Completers analysis set consisted of all subjects in the All Treated analysis set who completed the study. A subject was considered to have completed the study if he/she completed the six-month study visit with an observed value/non-missing SF-36 PCS at six months. This is an exclusive group of patients that does not include either discontinuers or switchers. All patients in this group have completed a six-month study visit with an observed value/non-missing SF-36 PCS at six months, independent of their responder/non-responder status based on SF-36 PCS. The tests of statistical significance were two-sided unless otherwise specified; any test resulting in $p<0.05$ was considered statistically significant. Two-sided 95% confidence intervals (CI) were used to assess the precision of endpoints, where relevant. Summaries and analyses were based on the All Treated analysis set unless otherwise specified. Subjects with an improvement from baseline in SF-36 PCS score of five points or above at six months from baseline were considered to be PCS responders. Those with no improvement or less than five points on the SF-36 PCS score were considered to be PCS non-responders. **Power and Sample Size:** To date, there is no information in literature for the estimation of appropriate patient numbers required for CART analysis. In general, CART analysis is an iterative process to arrive at an optimal decision tree with robust performance measures (predictive accuracy) and large datasets are recommended. A sample size of 950 patients was chosen as it allowed a precise estimation of the proportion of patients with improvement of five points on the PCS. It was also expected to be a large enough number to permit a CART analysis using several predictors with sufficient numbers of patients represented in the various nodes of the tree.

Table 1 Patient Demographic and Behavioral Characteristics at Baseline

Characteristics	Overall N=963
Age, mean (SD)	42.7 (12.9)
Gender, n (%)	
Male	591 (61.4%)
Female	372 (38.6%)
Race, n (%)	
White/ Caucasian	742 (77.1%)
Black/ African American	4 (0.4%)
Asian/ Middle Eastern	14 (1.5%)
Unknown	2 (0.2%)
Other	3 (0.3%)
Missing	198 (20.6%)
Education, n (%)	
None	10 (1.0%)
Primary school	83 (8.6%)
Secondary school	324 (33.6%)
High school	269 (27.9%)
University/post-university degree	193 (20.0%)
Unknown	84 (8.7%)
Work status, n (%)	
Full-time	520 (54.0%)
Part-time	79 (8.2%)
Unemployed	112 (11.6%)
Other	223 (23.2%)
Unknown	29 (3.0%)
Marital status, n (%)	
Married/ Common Law	592 (61.5%)
Divorced/ Separated	84 (8.7%)
Single, never married	228 (23.7%)
Widowed	12 (1.2%)
Unknown	47 (4.9%)
Smoking status, n (%)	
Never smoker	506 (52.5%)
Former smoker	158 (16.4%)
Current smoker	297 (30.8%)
Missing	2 (0.2%)
Drinking status, n (%)	
Never drinker	529 (54.9%)
Former drinker	133 (13.8%)
Current drinker	299 (31.0%)
Missing	2 (0.2%)
BMI (kg/m^2), mean (SD)	26.5 (4.8)

Table 2 Selected Patient Clinical Characteristics for the Study Population

Characteristics	Overall N=963
Symptom duration (years), Mean (SD)	11.6 (10.5)
Diagnosis duration (years), Mean (SD)	5.3 (7.7)
HLA-B27 Genotype	
Positive	614 (63.8%)
Negative	229 (23.8%)
Not evaluated	120 (12.5%)
Berlin Enthesitis Score	
Mean (SD)	1.8 (2.5)
Data not available	21 (2.2%)
Berlin Enthesitis Score (Category 1)	
n=0	447 (46.4%)
n>0	495 (51.4%)
Data not available	21 (2.2%)
Berlin Enthesitis Score (Category 2)	
≥ mean	414 (43.0%)
< mean	528 (54.8%)
Data not available	21 (2.2%)

Table 3 Reasons for Patient Discontinuation

	Number of patients discontinuing from study n (%) N=123
Lack of Efficacy	44 (35.8%)
Adverse Event(s)	32 (26.0%)
Lost to follow up	20 (16.3%)
Withdrew Consent	4 (3.3%)
Other:*	23 (18.7%)

*Other reasons include reasons linked to ineligibility for study, and inability to complete treatment or visits.

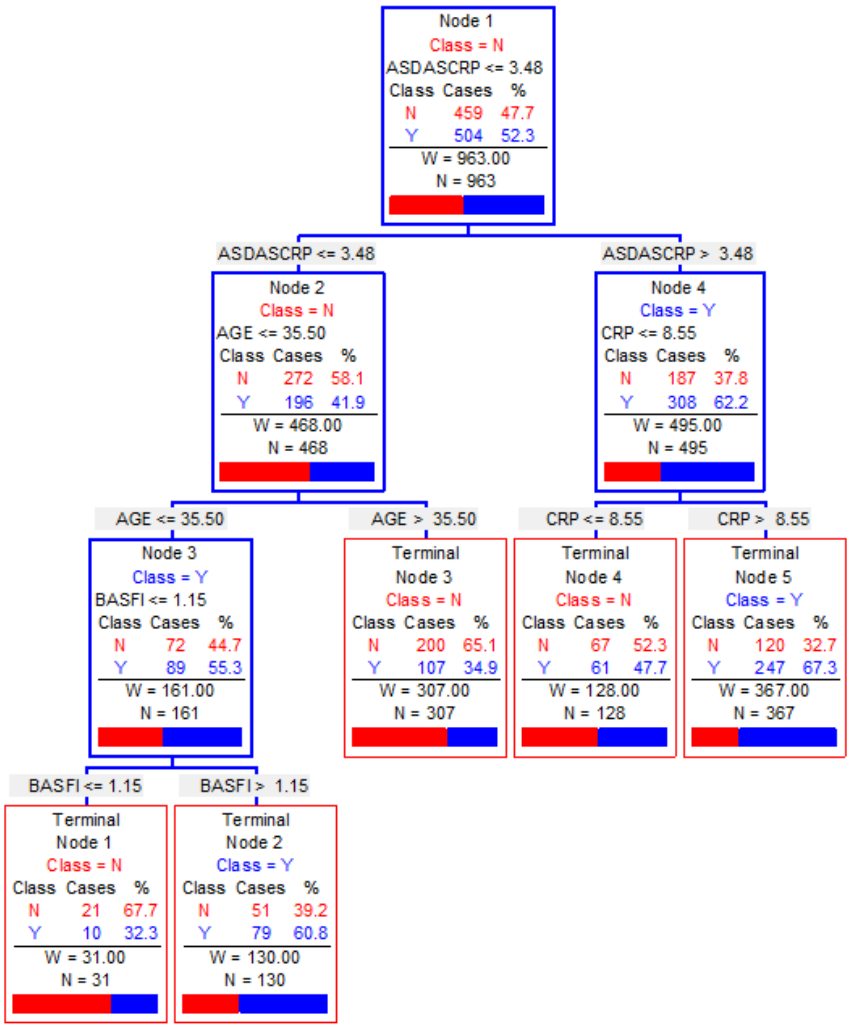
Analysis description	**Primary Efficacy Analysis**: Algorithm using baseline parameters that predicted change in HRQoL as measured by SF-36 PCS **Statistical methodology**: The CART modeling approach was used for developing the algorithm for the analysis of the primary endpoint. In this method, each explanatory variable was split into two parts, so called binary recursive partitioning, and the model evaluated all possible splits. Thus, the parent nodes were always split into exactly two child nodes. The term "recursive" indicates that the process was repeated by treating each child node as a parent. The selected predictive factors, root node and all subsequent child nodes were presented along with the corresponding binary partitioning cut off points. The final tree presented the explanatory variables with the greatest impact on the dependent variable. The dependent variable of the CART analysis was SF-36 PCS. Improvement of ≥ five points (≥ three points in a sensitivity analysis) was used to define PCS responders and PCS non-responders. The baseline predictor variables that were entered into the CART model were demographic (age, gender); clinical variables (symptom duration, HLA-B27 genotyping, enthesitis score, CRP); disease activity variables (BASDAI score, BASFI score, ASDAS score, PGA of disease, PGA of pain); and other (number of comorbidities at baseline < four or ≥ four).
Analysis population and time point description	All Treated analysis set; Timepoints – From baseline to six months
Summary	The mean change in SF-36 PCS at six months from baseline was 8.2 ± 8.4. At six months, 504 patients (52.3%) had an improvement in the SF-36 PCS of ≥ five points (PCS responders), and 444 patients (46.1%) had an improvement in the SF-36 PCS of < five points (PCS non-responders) (Table 4). Among the baseline variables considered for the most optimal CART tree, four were selected by CART for predicting PCS response at six months in this population. These are, in order of importance: **ASDAS** (categorized as ≤3.48, >3.48), **CRP** (categorized as ≤8.55, >8.55 mg/L), **age** (categorized as ≤35.50, >35.50 years) and **BASFI** (categorized as ≤1.15, >1.15) (Figure 1). Higher ASDAS and higher BASFI scores indicate higher disease activity. The first decision node was based upon ASDAS. Patients with ASDAS > 3.48 (n=495) had 62.2% PCS responders and were further split into two groups according to CRP (cutoff of 8.55 mg/L). Terminal nodes produced from this split led to 67.3% PCS responders among patients with a higher CRP (>8.55 mg/L) and 47.7% PCS responders among patients with a lower CRP (≤8.55 mg/L).

Summary (Cont.)	The population with ASDAS ≤ 3.48 was further split into two nodes based upon age (cutoff of 35.5 years). A higher proportion of PCS responders were identified in the age group less than or equal to 35.5 years compared to patients older than 35.5 years (55.3% vs. 34.9%, respectively). The younger group of patients was further split into two terminal nodes by the BASFI variable at an optimal cutoff of 1.15. A higher BASFI score was associated with more PCS responders at the end of study follow-up at six months compared to a BASFI score equal to or lower than 1.15. No further split was observed among patients older than 35.5 years. Based on the ten-fold cross-validation test sample, the CART tree correctly classified 57.5% of PCS responders (sensitivity) and 61.0% of PCS non-responders (specificity). The CRP variable produced the most optimal split, leading to the highest discrimination by PCS response in the study population. The BASFI variable produced the least discrimination by PCS response among the predictors identified by the CART procedure. The Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) for the test sample was 0.61 with a misclassification rate of 40.8%.

Table 4 Summary of SF-36 PCS Scores

SF-36 PCS Score	Overall N=963	Responder N=504	Non-responder N=459	P-value
Baseline	N=941	N=504	N=437	
Mean (SD)	34.7 (7.5)	33.5 (6.9)	36.1 (7.9)	<0.001
3 Months	N=885	N=496	N=389	
Mean (SD)	41.8 (9.2)	44.4 (8.4)	38.4 (9.0)	<0.001
6 Months	N=831	N=504	N=327	
Mean (SD)	43.0 (9.0)	46.8 (7.3)	37.2 (8.4)	<0.001
Change of PCS from Baseline				
3 Months Mean (SD)	N=876	N=496	N=380	
	7.0 (8.0)	10.9 (7.1)	2.0 (6.1)	<0.001
6 Months Mean (SD)	N=818	N=504	N=314	
	8.2 (8.4)	13.3 (6.2)	0.1 (4.2)	<0.001

Figure 1. Classification and Regression Tree Prediction of SF-36 PCS Response at Six Months among Patients Receiving at Least One Dose of Treatment



*Y= PCS responders, i.e., patients with PCS score improvement at six months ≥ five points; N=PCS nonresponders, i.e., patients with no PCS score improvement at six months or PCS score change < five points

Analysis description	**Secondary Efficacy analysis:** Alternative algorithm using baseline parameters that predicted change in HRQoL as measured by SF-36 MCS **Statistical methodology**: Similar analysis, including both CART analysis and step-wise multivariate logistic regression, as described for the primary endpoint was performed for SF-36 MCS response at six months. Improvement of ≥ five points (≥ three points in a sensitivity analysis) was used to define MCS responders and MCS non-responders.

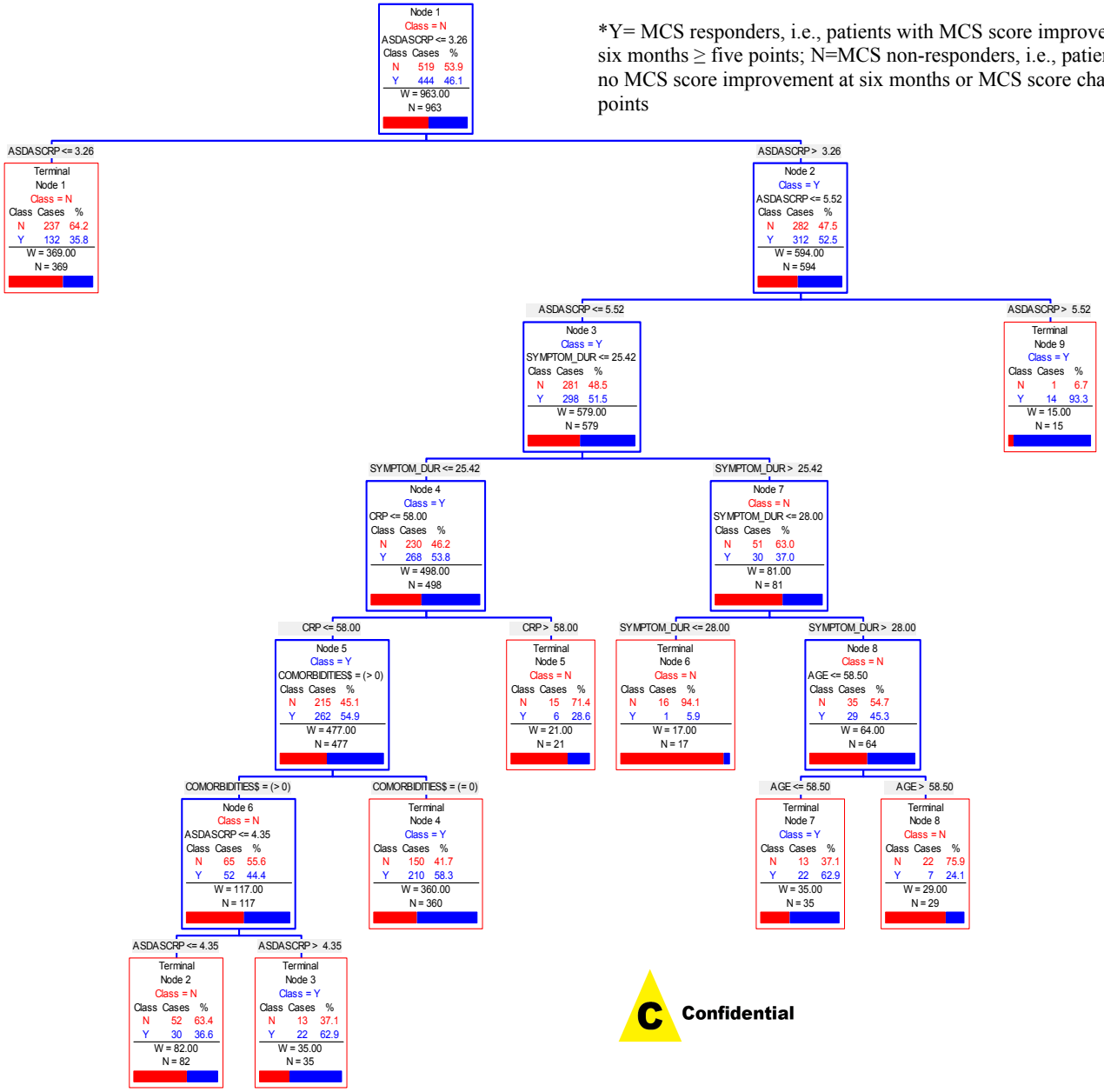
Analysis population and time point description	All Treated analysis set Timepoints – From baseline to six months
Summary	The mean ± SD SF-36 MCS score was significantly higher for MCS responders compared to MCS non-responders at baseline (41.1 ± 11.0 vs. 38.4 ± 11.1, respectively, p<0.001). The mean change of SF-36 MCS from baseline was significantly greater in MCS responders compared to non-responders at three months (7.0 ± 9.5 vs. 4.1 ± 9.9) and six months (7.6 ± 10.9 vs. 4.8 ± 10.0) (Table 5). Five baseline predictors were selected by CART for predicting MCS response (a change in MCS at six months from baseline with cutoff of five points) at six months in the total sample (N=963). These are, in order of importance: **ASDAS** (categorized as ≤3.26, >3.26 and as ≤5.52, >5.52), **symptom duration** (categorized as ≤25.42, >25.42 years and as ≤28.0, >28.0 years), **CRP** (categorized as ≤58.0, >58.0 mg/L), **age** (categorized as ≤58.5, >58.5 years) and **number of comorbidities** (categorized as =0, >0) (Figure 2). The first decision node was based upon ASDAS at an optimal cutoff of 3.26. Patients with ASDAS > 3.26 were further split into two nodes by ASDAS at higher cutoff of 5.52. No further split was observed among patients with ASDAS ≤ 3.26 and at the node with ASDAS > 5.52. The group of patients with ASDAS ≤ 5.52 was split by the symptom duration variable at a cutoff of 25.42 years. Patients with a higher symptom duration (>25.42 years) had a lower proportion of responders compared to patients with a lower symptom duration (≤25.42 years) (37.0% vs. 53.8%, respectively), and were further split by the symptom duration variable into two groups at a cutoff of 28. No further split was observed among patients with symptom duration of ≤28. The age variable split patients with symptom duration of >28 into two nodes at a cutoff of 58.5 years. Older patients had a lower proportion of responders compared to younger patients (24.1% vs. 62.9%, respectively) although based on a small sample. Patients with a lower CRP (≤58.0 mg/L) were further split into two nodes by the number of comorbidities variable yielding 58.3% responders among patients with no comorbidities and 44.4% responders among patients with at least one comorbidity. Patients with a comorbidity were further split into two groups by the ASDAS variable at an optimal cutoff of the ASDAS variable at an optimal cutoff of 4.35 yielding a higher proportion of responders in the group with ASDAS > 4.35 compared to ASDAS ≤ 4.35 (62.9% vs. 36.6%, respectively). Based on the ten-fold cross-validation test sample, the CART tree correctly classified 48.0% of

Summary (Cont.)	MCS responders (sensitivity) and 61.9% (specificity) of MCS non-responders. The ROC-AUC for the test sample was 0.54 with a misclassification rate of 44.6%.

Table 5 Summary of SF-36 MCS Score

SF-36 MCS Score	Overall N=963	Responder N=504	Non-responder N=459	P-value
Baseline	N=941	N=504	N=437	
Mean (SD)	39.8 (11.1)	41.1 (11.0)	38.4 (11.1)	<0.001
3 Months	N=885	N=496	N=389	
Mean (SD)	45.6 (10.4)	48.0 (9.5)	42.6 (10.8)	<0.001
6 Months	N=831	N=504	N=327	
Mean (SD)	46.3 (10.6)	48.7 (10.0)	42.6 (10.4)	<0.001
Change of MCS from Baseline				
3 Months Mean (SD)	N=876	N=496	N=380	
	5.8 (9.8)	7.0 (9.5)	4.1 (9.9)	<0.001
6 Months Mean (SD)	N=818	N=504	N=314	
	6.5 (10.6)	7.6 (10.9)	4.8 (10.0)	<0.001

Figure 2 - Classification and Regression Tree Prediction of SF-36 MCS Response at Six Months among Patients Receiving at Least One Dose of Treatment



*Y= MCS responders, i.e., patients with MCS score improvement at six months ≥ five points; N=MCS non-responders, i.e., patients with no MCS score improvement at six months or MCS score change < five points

Analysis description	**Secondary Efficacy Endpoint:** Association beween change in Disease Severity and change in HRQoL as measured by SF-36 PCS **Statistical methodology:** Change in BASDAI score and BASDAI50 response was used as the main measures of disease severity. In addition, ASDAS and ASAS responses were investigated as alternative measures. Summaries of BASDAI and ASDAS change from baseline at three months and six months, as well as BASDAI50, ASDAS clinically important improvement, ASDAS major improvement, ASAS20/ASAS40 response and ASAS partial remission at three months and six months were presented stratified by SF-36 PCS responders and PCS non-responders using the cutoff point of five at six months. As statistical tests, a two-sample t-test was used for continuous variables, and a Pearson Chi square test was used for categorical variables.
Analysis population and time point description	All Treated analysis set Timepoints – From baseline to three months and six months.
Summary	**BASDAI**: The mean ± SD BASDAI score for all patients at baseline was 6.2 ± 1.9. The change from baseline at three months and six months was -2.4 ± 2.2 and -2.7 ± 2.3, respectively. The mean ± SD BASDAI score for PCS responders was 6.3 ± 1.8 and for PCS non-responders was 6.1 ± 2.0 at baseline. The mean ± SD reduction in BASDAI score from baseline was significantly greater among PCS responders compared to PCS non-responders at three months (-3.1 ± 2.0 vs. -1.4 ± 2.0, respectively, p<0.001) and at six months (-3.6 ± 2.0 vs. -1.3 ± 2.0, respectively, p<0.001). The overall proportion of patients achieving BASDAI50 response at three months and six months was 34.6% (n=333) and 39.5% (n=380). The proportion of PCS responders achieving BASDAI50 criteria was significantly higher compared to PCS non-responders at three months (49.2% vs. 18.5%, p<0.001) and at six months (64.5% vs. 12.0%, respectively, p<0.001). **BASFI**: The mean ± SD BASFI score for all patients at baseline was 5.3 ± 2.4. The change from baseline at three months and six months was -1.8 ± 2.1 and -2.1 ± 2.3, respectively. The mean ± SD BASFI score at baseline for PCS responders was 5.5 ± 2.3 and for PCS non-responders was 5.2 ± 2.6. The mean ± SD reduction in BASFI score from baseline was significantly greater among PCS responders compared to PCS non-responders at three months (-2.6 ± 2.0 vs. -0.9 ± 1.9, respectively, p<0.001) and at six months (-3.0 ± 2.1 vs. -0.8 ± 1.9, respectively, p<0.001). **ASDAS**: ASDAS major improvement was defined as decrease from baseline of ≥ 2.0 units. ASDAS clinically important improvement was defined as decrease from baseline of ≥ 1.1 units. The mean ± SD ASDAS-CRP at baseline, three months and six months was 3.6 ± 1.0, 2.2 ± 1.0, and

Summary (Cont.)	2.1 ± 1.0, respectively. Overall, 26.6% (n=256) of patients achieved ASDAS major improvement and 47.9% (n=461) achieved ASDAS clinically important improvement at six months. The mean ± SD ASDAS score for PCS responders was 3.7 ± 1.0 and for PCS non-responders was 3.4 ± 0.9 at baseline. The mean ± SD ASDAS score was lower among PCS responders compared to PCS non-responders at three months (2.0 ± 1.0 vs. 2.5 ± 1.0, respectively) and at six months (1.7 ± 0.9 vs. 2.6 ± 1.0). At three months, PCS responders and PCS non-responders were classified based on the ASDAS score categories into inactive (PCS responders: n=134, 26.6%; PCS non-responders: n=40, 8.7%), moderate (PCS responders: n=127, 25.2%; PCS non-responders: n=76, 16.6%), high (PCS responders: n=153, 30.4%; PCS non-responders: n=167, 36.4%), and very high (PCS responders: n=40, 7.9%; PCS non-responders: n=64, 13.9%). At six months, PCS responders and PCS non-responders were classified based on the ASDAS score categories into inactive (PCS responders: n=157, 31.2%; PCS non-responders: n=35, 7.6%), moderate (PCS responders: n=146, 29.0%; PCS non-responders: n=58, 12.6%), high (PCS responders: n=128, 25.4%; PCS non-responders: n=155, 33.8%), and very high (PCS responders: n=16, 3.2%; PCS non-responders: n=51, 11.1%). A significantly higher proportion of PCS responders achieved ASDAS major improvement compared to PCS non-responders at three months (38.3% vs. 8.3%, respectively, p<0.001) and at six months (45.0% vs. 6.3%, respectively, p<0.001). A significantly higher proportion of PCS responders also achieved ASDAS clinically important improvement at three months (62.5% vs. 27.2%, respectively, p<0.001) and at six months (71.6% vs. 21.8%, respectively, p<0.001). **PGA of Disease:** The mean ± SD PGA of disease activity score at baseline for all patients was 6.6 ± 2.3. The change from baseline at three months and six months was -2.5 ± 2.8 and -2.8 ± 2.9, respectively. The mean ± SD PGA of disease activity score for PCS responders was (6.8 ± 2.2) and for PCS non-responders (6.4 ± 2.5) at baseline. The mean ± SD reduction in PGA of disease activity from baseline was significantly greater for PCS responders at three months (-3.4 ± 2.6 vs. -1.3 ± 2.6, respectively, p<0.001) and at six months (-3.8 ± 2.7 vs. -1.2 ± 2.6, respectively, p<0.001). **PGA of Pain:** The mean ± SD PGA of pain score at baseline for all patients was 6.7 ± 2.3. The change from baseline at three months and six months was -2.5 ± 2.8 and -3.0 ± 2.8, respectively. The mean ± SD PGA of pain score for PCS responders was 6.8 ± 2.2 and for PCS non-responders was 6.5 ± 2.4 at baseline. The reduction in PGA of pain from baseline was significantly greater for PCS responders compared to PCS non-responders at three months (-3.4 ± 2.5 vs. -1.4 ± 2.6, respectively, p<0.001) and at six months (-4.0 ± 2.5 vs. -1.5 ± 2.5, respectively, p<0.001).

Summary (Cont.)	**ASAS**: At three months, the overall proportion of patients achieving ASAS20 response, ASAS40 response, and ASAS partial remission was 48.5% (n=467), 29.7% (n=286) and 12.1% (n=117). At six months, the overall proportion of patients achieving ASAS20 response, ASAS40 response and ASAS partial remission was 50.6% (n=487), 34.6% (n=334), and 14.3% (n=138). A significantly higher proportion of PCS responders achieved ASAS20 response compared to PCS non-responders at three months (70.2% vs. 24.6%, respectively. p<0.001) and at six months (77.6% vs. 20.9%, p<.001). Similarly, a significantly higher proportion of PCS responders achieved ASAS40 criteria compared to PCS non-responders at three months (44.8% vs. 13.1%, respectively, p<0.001) and at six months 59.1% vs. 7.8%, p<0.001). The proportion of PCS responders achieving partial remission was significantly higher compared to that of non-responders at three months (19.0% vs. 4.6%, respectively, p<0.001) and at six months (23.4% vs. 4.4%, p<0.001).
Analysis description	**Secondary Efficacy Endpoint:** Health Care Resource Utilization and Work Productivity **Statistical methodology:** Resource utilization at baseline represents the AS related resource utilization during the three months prior to initiation of anti-TNF treatment with either infliximab or golimumab. All descriptive statistics for health resource utilization and work productivity were presented by overall, by PCS responder/non-responder (≥ five points of improvement in PCS at six months). Statistical significance of the difference between PCS responders and PCS non-responders was assessed using appropriate statistical tests. Pearson Chi square was used for categorical data and Student's t-tests for continuous data. For categorical data specified to be analyzed using the Pearson Chi square test, Fisher's exact test was used if the Pearson Chi square test was inappropriate.
Analysis population and time point description	All Treated analysis set Timepoints – From baseline to three months and six months.
Summary	Health care resource utilization decreased from baseline at six months for the overall study population with a decrease in inpatient, acute care and outpatient visits. No significant diffrences were observed between PCS responders and PCS non-responders at baseline and at six months in health care resource utilization

Summary (Cont.)	**Acute Care**: At baseline, 1.6% (n=15) patients received acute emergency care. At six months, overall, 0.3% (n=3) patients received acute care. At baseline, six PCS responders and nine PCS non-responders received acute emergency care. At six months, overall, 0.3% (n=3) patients received acute care of whom two were PCS responders and one was a PCS non-responder. **Inpatient Care**: At baseline, 13.6% (n=131) of the patients received inpatient care; at six months, 3.1% (n=30) of the patients received inpatient care. No significant differences were observed between PCS responders and PCS non-responders at baseline and at six months. At baseline, 14.9% (n=75) of the PCS responders and 12.2% (n=56) of the PCS non-responders received inpatient care. At six months, 2.8% (n=14) of PCS responders and 3.5% (n=16) of PCS non-responders received inpatient care. **Outpatient Care:** At baseline, 39.4% (n=379) of the patients received outpatient/day care. The mean ± SD number of visits was 2.4 ± 1.8. At six months, 19.0% (n=183) of the patients received outpatient/day care. The mean ± SD number of outpatient/day care visits was 2.1 ± 2.7. No significant differences were observed between PCS responders and PCS non-responders at baseline and at six months. At baseline, the proportion of PCS responders receiving outpatient/day care was 38.9% and the proportion of PCS non-responders receiving outpatient/day care was 39.9%. At six months, the proportion of PCS responders receiving outpatient/day care was 19.4% and the proportion of PCS non-responders receiving outpatient/day care was 18.5%. **Concomitant Medication**: Overall, 84.3% (n=812) patients used concomitant medications for AS treatment. The mean ± SD duration of concomitant medication use among all patients was 353.4 ± 237.9 days. Among the patients using concomitant medications, the majority used NSAIDs (n=663, 68.8%); followed by DMARDs (n=239, 24.8%), corticosteroids (n=128, 13.3%), analgesics (n=247, 25.6%), and medication for extra-articular manifestations (n=65, 6.7%). A lower proportion of PCS responders (n=412, 81.7%) used concomitant medications for AS treatment compared to PCS non-responders (n=400, 87.1%). The mean ± SD duration of concomitant medication use was similar between PCS responders and PCS non-responders (349.6 ± 245.1 vs. 357.2 ± 230.6, p=0.662). The distribution of patients by concomitant medication type was similar between PCS responders and non-responders. **Work Productivity and Activity Impairment**: Overall, the mean number of work days missed due to AS at baseline was 6.3 ± 31.1 and this decreased to 2.7 ± 12.3 at six months. The mean ± SD work impairment score was 48.3 ± 34.0 at baseline and the change from baseline at six months was -24.3 ± 34.8. The mean ± SD activity impairment score was

Summary (Cont.)	62.5 ± 24.7 at baseline and the change from baseline at six months was -26.9 ± 28.5. The mean ± SD presenteeism score was 51.9 ± 30.6 at baseline and the change from baseline at six months was -23.6 ± 31.9. The mean ± SD absenteeism score was 16.4 ± 32.3 at baseline and the change from baseline at six months was -8.7 ± 31.5. There were no significant differences in the mean number of work days missed due to AS at baseline between PCS responders and PCS non-responders. The mean ± SD change in work impairment scores was significantly greater among PCS responders compared to PCS non-responders (-31.8 ± 32.4 vs. -8.6 ± 34.4, p<0.001) from baseline to six months. The mean ± SD change in activity impairment scores was significantly greater among PCS responders compared to PCS non-responders (-37.7 ± 25.7 vs. -10.2 ± 24.3, p<0.001) from baseline to six months. The mean ± SD change in presenteeism scores was significantly greater among PCS responders compared to PCS non-responders (-32.2 ± 29.3 vs. -5.4 ± 29.6, p<0.001) from baseline to six months. The mean ± SD change in absenteeism scores was significantly greater among PCS responders compared to PCS non-responders (-11.5 ± 29.1 vs. -2.9 ± 35.3, p=0.006) from baseline to six months.
Analysis description	**Secondary Efficacy Endpoint:** External Validity of Clinical Algorithm to identify patients who responded to AS therapy **Statistical methodology**: To investigate the external validity of a clinical algorithm for identifying patients who responded to AS therapy, multivariate logistic regression analyses was performed for the outcome variables BASDAI50 response, ASDAS clinically important improvement, ASDAS major improvement, ASDAS inactive disease, ASAS20 response, and ASAS partial remission at six months. Only the variables identified as predictors in the paper by Vastesaeger et al[1] were included in the multivariate logistic regression model. The predictor variables at baseline were categorized according to the paper. • Age: >40 vs. ≤40 • HLA-B27 genotype: positive vs. negative • BASFI score: >6.5, >4.5 and ≤6.5, ≤4.5 • Berlin enthesitis score: =0 vs. >0 • CRP: <0.6, ≥0.6 and ≤2.0, >2.0 To assess the performance of the logistic regression models used in this study compared to the model developed by Vastesaeger et al, the ROC-AUC and Hosmer-Lemeshow test result were presented. Matrix representations of response rate for each outcome variable were also presented.

Analysis population and time point description	Timepoints – From baseline to six months.
Summary	Table 6 shows results from multivariate logistic regression models for outcome variables BASDAI50 response, ASDAS clinically important improvement, ASDAS major improvement, ASDAS inactive disease, ASAS20 response, and ASAS partial remission at six months. The predictors in the models were based on variables identified as predictors in the paper by Vastesaeger et al.[1] ROC-AUC values presented in the table, ranging from 0.65 to 0.78, represent the performance of these models compared to the model developed by Vastesaeger et al.[1] Age older than 40 years, HLA-B27 positive genotype, and CRP greater than 6 mg/L were associated with higher odds of ASDAS clinically important improvement and ASDAS major improvement in two separate models. Age older than 40 years, HLA-B27 positive genotype, BASFI greater than 6.5, and CRP greater than 6 mg/L were associated with higher odds of ASDAS inactive disease, BASDAI-50 response, ASAS20 response, and ASAS partial remission in four separate models.

Table 6: External Validity of a Clinical Algorithm to Predict AS Response

Baseline Characteristics	Odds Ratio (95% CI)	ROC-AUC*
Outcome 1: ASDAS Clinically Important Improvement		0.76
Age (≤40 vs. >40)	1.7 (1.2, 2.4)	
HLA-B27 Genotype (Positive vs. Negative)	2.4 (1.7, 3.5)	
BASFI (≤4.5 vs. >6.5)	0.8 (0.5, 1.2)	
(>4.5 and ≤6.5 vs. >6.5)	0.9 (0.6, 1.3)	
Berlin Enthesitis score (=0 vs. >0)	1.0 (0.7, 1.5)	
CRP (≥6 and ≤20 vs. <6)	2.8 (1.9, 4.0)	
(>20 vs. <6)	9.0 (5.4, 15.0)	
Outcome 2: ASDAS Major Improvement		0.78
Age (≤40 vs. >40)	1.6 (1.1, 2.3)	
HLA-B27 Genotype (Positive vs. Negative)	2.0 (1.3, 3.1)	
BASFI (≤4.5 vs. >6.5)	0.7 (0.4, 1.0)	
(>4.5 and ≤6.5 vs. >6.5)	1.2 (0.8, 1.9)	
Berlin Enthesitis score (=0 vs. >0)	1.1 (0.8, 1.5)	
CRP (≥6 and ≤20 vs. <6)	4.5 (2.9, 7.0)	
(>20 vs. <6)	11.1 (6.9, 17.8)	
Outcome 3: ASDAS Inactive Disease		0.71
Age (≤40 vs. >40)	1.5 (1.0, 2.1)	
HLA-B27 Genotype (Positive vs. Negative)	2.1 (1.4, 3.3)	
BASFI (≤4.5 vs. >6.5)	4.5 (2.8, 7.1)	
(>4.5 and ≤6.5 vs. >6.5)	2.9 (1.7, 4.7)	
Berlin Enthesitis score (=0 vs. >0)	1.1 (0.8, 1.5)	
CRP (≥6 and ≤20 vs. <6)	1.2 (0.8, 1.7)	
(>20 vs. <6)	0.8 (0.5, 1.3)	
Outcome 4: BASDAI-50 Response		0.68
Age (≤40 vs. >40)	1.7 (1.2, 2.3)	
HLA-B27 Genotype (Positive vs. Negative)	2.2 (1.5, 3.1)	
BASFI (≤4.5 vs. >6.5)	1.6 (1.1, 2.3)	
(>4.5 and ≤6.5 vs. >6.5)	1.3 (0.9, 1.9)	
Berlin Enthesitis score (=0 vs. >0)	1.0 (0.7, 1.4)	
CRP (≥6 and ≤20 vs. <6)	1.8 (1.3, 2.5)	
(>20 vs. <6)	3.1 (2.1, 4.6)	
Outcome 5: ASAS20 Response		0.65
Age (≤40 vs. >40)	1.6 (1.2, 2.2)	
HLA-B27 Genotype (Positive vs. Negative)	1.8 (1.3, 2.6)	
BASFI (≤4.5 vs. >6.5)	0.6 (0.4, 0.8)	
(>4.5 and ≤6.5 vs. >6.5)	0.9 (0.6, 1.3)	
Berlin Enthesitis score (=0 vs. >0)	0.9 (0.6, 1.2)	
CRP (≥6 and ≤20 vs. <6)	1.6 (1.2, 2.3)	
(>20 vs. <6)	2.0 (1.3, 2.9)	
Outcome 6: ASAS Partial Remission		0.76
Age (≤40 vs. >40)	1.8 (1.2, 2.7)	
HLA-B27 Genotype (Positive vs. Negative)	2.9 (1.6, 5.3)	
BASFI (≤4.5 vs. >6.5)	6.9 (3.7, 12.7)	
(>4.5 and ≤6.5 vs. >6.5)	3.4 (1.8, 6.7)	
Berlin Enthesitis score (=0 vs. >0)	1.2 (0.8, 1.8)	
CRP (≥6 and ≤20 vs. <6)	1.6 (1.0, 2.6)	
(>20 vs. <6)	1.9 (1.1, 3.1)	

*ROC: Receiver Operating Characteristic; AUC: Area Under the Curve

Analysis description	**Adverse Events** **Statistical methodology**:Spontaneously reported AEs and SAEs were tabulated by system organ class and preferred term.
Analysis population and time point description	Timepoints – From baseline to six months.
Summary	Throughout the six months follow-up, 22.1% (n=213) patients had at least one spontaneously reported AE. The most common AEs were infections and infestations, 7.7% (n=74), followed by general disorders and administration site conditions, 7.6% (n=73). Throughout the six months follow-up, 1.8% (n=17) patients reported a SAE. The most commonly reported SAEs were musculoskeletal and connective tissue disorders, 0.6% (n=6).

CONCLUSIONS:	The disease activity of the study population improved over six months of treatment as observed using the BASDAI and the ASDAS scores. Improvements were observed in HRQoL as measured by changes from baseline to six months in SF-36 PCS and MCS scores. Over 50% of the study population had an improvement in SF-36 PCS by five points or more. Patients with greater improvement in SF-36 PCS and MCS had improved disease activity, work productivity and activity impairment scores at six months. Although the CART models had moderate levels of sensitivity and specificity, the results show that baseline parameters indicating high disease activity and inflammation, such as high ASDAS score and slightly elevated CRP, combined with younger age, are associated with SF-36 PCS response. These results can provide insight to clinicians as to which patients will have a higher likelihood to achieve HRQoL improvement when being treated with infliximab or golimumab.
REFERENCE:	1. Vastesaeger N, van der Heijde D, Inman RD et al. Predicting the outcome of ankylosing spondylitis therapy. Ann Rheum Dis 2011;70(6):973-981.
REPORT DATE:	8-Dec-2015