Validation Exercise of the Ankylosing Spondylitis Assessment Study (ASAS) Group Response Criteria in Ankylosing Spondylitis Patients Treated With Biologics

MILICENT A. STONE,1 ROBERT D. INMAN,2 JAMES G. WRIGHT,3 AND ANDREAS MAETZEL1

Objectives. To define what expert clinicians consider a dramatic response in ankylosing spondylitis (AS) patients treated with biologic therapies based on patient and physician assessments of global disease activity. To compare this expert clinician-derived criteria to the Ankylosing Spondylitis Assessment Study (ASAS) Group criteria for improvement.

Methods. Forty consecutive AS patients were treated in a 1-year open-label protocol with infliximab. Response to treatment at week 52 was defined using ASAS response criteria. For the purpose of this exercise, improvement using ASAS criteria was defined by consensus among experts as good with 50% improvement from baseline (ASAS50) and dramatic with 70% improvement from baseline (ASAS70). Experts established separate criteria for improvement in disease activity as good or dramatic based on patient and physician global assessment of disease activity.

Results. Twelve of 40 patients met the ASAS 70 criteria, however, only 8 met the expert definition of a dramatic improvement based on physician global scores and 5 met the expert definition of a dramatic improvement based on patient global assessment of disease activity. Agreement was poor between ASAS50 or ASAS70 and expert definition of improvement based on physician global scores (k < 0.3), but agreement was moderate to good between ASAS50 or ASAS70 and expert definition of improvement based on patient global scores (k = 0.6–0.7).

Conclusions. Differential response experienced by AS patients treated with infliximab was adequately captured by the ASAS composite improvement criteria. Overall, this study demonstrates the validity of the ASAS criteria for the detection of improvement in AS patients treated with biologics. However, the patient global assessment of disease activity may be sufficient to monitor changes in disease activity in these patients.

KEY WORDS. Ankylosing spondylitis; Infliximab; Outcomes; ASAS; Relative improvement criteria; Validation

INTRODUCTION

Ankylosing spondylitis (AS), a chronic rheumatic disease of young adults, is the prototype of the spondyloarthropathy family of inflammatory arthropathies affecting up to 1.9% of the population (1). Spinal inflammation, the hallmark of AS, causes pain and stiffness leading to progressive spinal deformity and fusion from syndesmophytes (bony overgrowths in the spine). Many patients with AS face challenges in the workplace and approximately one-third leave the labor force within 20 years of diagnosis (2).

The traditional approach to treatment of AS has been the use of exercise and nonsteroidal antiinflammatory drugs (NSAIDs). Although these treatment modalities may reduce pain, neither has resulted in substantial reductions in disease activity or improvement in functional disability (3,4). Several recent short-term, randomized controlled trials have demonstrated tumor necrosis factor (TNF) recep-
tor blocking agents to be highly efficacious in patients with AS (5,6). The TNF receptor blockers are members of the family of biologic agents targeting the specific biological pathway of TNFα.

Improvement criteria for assessing response to biologics was recently proposed by the Ankylosing Spondylitis Assessment Study (ASAS) Group (7). Alternative definitions of improvement have been proposed by Brandt et al based on the results of a randomized controlled trial of infliximab in AS (6). Other criteria, such as absolute disease activity criteria (ASAS Workshop, Berlin 2003), attempt to differentiate a clinically important from a clinically unimportant response. Accounting for this shift is the recognition that a clinically important response would need to be separated from one that is clinically unimportant. A definition of what constitutes the smallest detectable difference in response to therapy targeted toward AS was recently proposed (9). Biologic agents can lead to a dramatic improvement in patients’ symptoms with marked reduction in disease activity and spinal inflammation. This dramatic response may not be adequately captured by the use of relative improvement criteria, nor by assessment of the smallest detectable difference (7,9).

The goal of this study was to elucidate what therapeutic response would qualify as dramatic response to biologics. The objectives were twofold: To define what expert clinicians constitute as a dramatic response in AS patients treated with biologic therapies based on patient and physician assessments of global disease activity; and to compare this expert clinician-derived criteria to the ASAS criteria for improvement.

PATIENTS AND METHODS

Patient population. Forty consecutive patients with AS meeting the modified New York criteria (10) were treated in an open-label protocol with infliximab, a monoclonal anti-TNFα antibody. All patients received 3 intravenous infusions at a dosage of 5 mg/kg over the first 6 weeks and a maintenance schedule every 8 weeks thereafter for 1 year. Patients were required to have tried conventional therapy without significant improvement and to meet 2 of the 3 following inclusion criteria on 2 separate occasions in the month preceding therapy: 1) Bath Ankylosing Spondylitis Disease Activity index (BASDAI) (11) score >4 (0 = no disease activity, 10 = maximum disease activity); 2) total body or spinal pain >2 (0 = no pain, 3 = maximum pain); and 3) early morning stiffness >30 minutes. Permission for the study was obtained from the hospital’s ethics committee, and written informed consent was obtained from all patients. Patients were allowed to continue taking their current disease-modifying antirheumatic drug (DMARD), if any, but were required to have been on a stable dosage for at least 1 month prior to study entry. Prednisone was allowed at a stable dose <10 mg/day for at least 1 month prior to entry. Patients were required to maintain the same dosage of NSAIDs during the treatment period.

Demographic and clinical data. The following disease characteristics were recorded at baseline: disease duration, pattern of disease (axial, peripheral manifestations, or both), history of anterior uveitis or other extraarticular manifestations of disease, prior drug therapy and concurrent DMARD or steroid therapy. The following measures were obtained at baseline and each followup visit: patient global assessment of disease activity (5-point Likert scale in which 0 = no disease activity, 4 = very severe disease activity); patient pain (visual analog scale [VAS]) in which 0 = none, 100 = maximum); Bath Ankylosing Spondylitis Functional Index score (BASFI in which 0 = none, 100 = worst); early morning stiffness (calculated as a mean of the following 2 scales: VAS in which 0 = no stiffness, 100 = maximum average early morning stiffness in the preceding week; and VAS with a 100-mm categorical scale of stiffness ranging from 0 to 120 minutes in 30-minute increments); and physician global assessment of disease activity (4-point Likert scale in which 0 = no disease activity, 1 = very severe disease activity).

Measures of improvement. ASAS criteria. Response to treatment was defined using the ASAS criteria (7). The definition of ASAS improvement is ≥20% relative improvement and absolute improvement of ≥10 units in 3 or more of the following 4 domains: inflammation (mean of questions 5 and 6 of the BASDAI), function (BASFI), patient perception of pain (patient pain VAS), and patient global assessment, with no worsening in the fourth domain. We defined an ASAS50 and ASAS70 response as attaining a 50% and 70% relative improvement, respectively, and an absolute improvement of ≥20 units in 3 of the 4 domains from baseline to study end with no worsening permitted in the fourth domain. Experts agreed by consensus that attaining an ASAS50 should be considered a good improvement and, similarly, attaining an ASAS70 should be considered a dramatic improvement for the purpose of this validation exercise.

Expert definition of response. A survey was undertaken among 10 Canadian rheumatologists, experts in the management of spondylitis. This survey was conducted at a focus group of the Spondyloarthropathy Research Consortium of Canada (SPARCC) working party. All SPARCC attendees participated and were asked to comment on 2 questions pertaining to clinical scenarios of AS patients treated with infliximab for 1 year. In the first part of the exercise, we asked respondents to classify all possible improvements on the 5-point Likert scale of patient global disease activity into dramatic, good, moderate, mild, or no improvement. In the second part of the exercise, we asked respondents to classify all possible improvements on the 4-point Likert scale of physician global assessment into dramatic, good, moderate, mild, or no improvement.

Patient definition of dramatic response. To investigate if there was a discrepancy between the patients’ and physicians’ perception of dramatic improvement to a biologic agent, we asked 15 consecutive AS patients who had received at least 1 year of treatment with a biologic agent to characterize what they would consider a dramatic improvement in disease activity. The question was framed in
a similar manner to that presented to the physicians when they were asked to comment on shifts in patient global scores. Patients were asked to classify all possible improvements on the 5-point patient global Likert scale into dramatic, good, moderate, mild, or no improvement.

**Analysis.** All 40 patients based on their response between baseline and study end were classified by ASAS criteria using ASAS50 and ASAS70 cutoffs. All 40 patients in the study were also classified using the expert definition of improvement based on global patient and physician scores as having a dramatic improvement or not, or a good improvement or not. Cohen’s kappa agreement statistics were calculated between expert definitions of improvement and the ASAS20, ASAS50, and ASAS70 criteria for improvement (13). Agreement that was <50% beyond chance alone (κ < 0.5) was considered poor, between 50% and 70% (κ = 0.5–0.7) was considered moderate, and (κ > 0.7) was considered good.

**RESULTS**

Forty patients were enrolled with a male:female ratio of 33:7, a mean ± SD age of 38 ± 9 years, and a mean ± SD disease duration of 10 ± 7 years. At study entry, 10 patients were taking prednisone, 25 were taking DMARDs, and 34 were taking NSAIDs. All 40 patients completed the induction phase of the protocol. Nine patients discontinued drug before year 1 for the following reasons: 4 because of lack of efficacy (2 at 22 weeks and 1 at 6 weeks), 3 because of adverse effects (all at 46 weeks), and 2 because they were unable to afford it. Thirty-seven patients were available at 46 weeks for followup and 31, at 1 year. In these subjects, the last observation point for all outcomes was carried forward.

Overall there was improvement in the patient global assessment of disease activity with infliximab over the 52-week period (Figure 1).

**Expert definition of improvement.** Of the 10 experts surveyed, 1 did not return the questionnaire; therefore, the results are based on the ratings of 9 experts. Experts almost unanimously agreed that shifts in disease activity from severe or very severe to no disease activity by patient global assessment or from severe to no disease activity by physician global assessment would qualify as dramatic improvement (Table 1). There was less agreement on how to define shifts from severe or very severe to mild, with some qualifying these as good improvements and others as dramatic. We therefore decided to classify these in the good improvement category.

There was 100% agreement among the 15 AS patients surveyed that an improvement from very severe or severe to no disease activity constituted a dramatic improvement. Therefore, the expert definition of improvement based on patient global assessment scores was considered an accurate reflection of the patients’ perception of disease activity.

**Dramatic improvement.** Twelve (30%) of 40 patients met ASAS70 response criteria, but only 8 (20%) and 5 (13%) patients had dramatic improvement by expert criteria based on physician and patient global scores, respectively (13). Agreement that was <50% beyond chance alone (κ < 0.5) was considered poor, between 50% and 70% (κ = 0.5–0.7) was considered moderate, and (κ > 0.7) was considered good.

**Good improvement.** Twenty-two (54%) patients met ASAS50 response criteria as compared with 21 (52%) by expert definitions of improvement based on physician global scores and 18 (45%) by patient global scores. Agreement was poor between ASAS50 and expert definitions for dramatic improvement based on physician global scores (κ < 0.3), but agreement was moderate for expert definition for improvement based on patient global scores (κ = 0.6; Table 2). These discrepancies largely resulted from the fact that 4 of the 8 patients who were considered as having dramatic improvement by the expert definition based on physician global scores did not fulfill ASAS70 criteria. Approximately two-thirds (8 and 7) of the 12 patients who were classified as ASAS70 responders were not considered dramatic responders per the expert definition based on physician or patient global assessment of disease activity (Table 2). Interestingly, all patients who fulfilled the ASAS70 improvement criteria also met the ASAS definition for partial remission defined as an end-of-trial value <20/100 in each of the 4 domains (7).

**Figure 1.** Patients’ global assessment of change in disease activity, on a 5-point Likert scale, between baseline and study end or discontinuation of the study drug. The blue lines represent the remaining patients (n = 28) who did not meet the ASAS50 definition for improvement. The green lines represent those patients (n = 12) who met ASAS70 criteria for improvement. The black lines represent those patients (n = 12) who met ASAS70 criteria for improvement. The black lines represent those patients (n = 12) who met ASAS70 criteria for improvement.
DISCUSSION

The long-term safety and efficacy profiles of biologic agents in AS are as yet unknown. These uncertainties warrant a more careful weighing of decisions, such as who should be treated and when treatment should be stopped. Furthermore, the high costs of biologic agents make it imperative that physicians gain an understanding of the degree to which patients benefit from these new therapies. While relative improvement criteria have been criticized for not adequately conveying the dramatic improvements in disease activity observed with biologic agents in the rheumatic diseases (14,15), our results corroborate the use of relative improvement cutoffs of 50% relative improvement or 70% relative improvement in the ASAS criteria because they do not miss what experts consider dramatic shifts in the patient’s global assessment of disease activity. A potential explanation for this is that the ASAS composite improvement criteria, unlike the American College of Rheumatology relative improvement criteria for rheumatoid arthritis (16), require both relative and absolute improvement and may therefore more accurately capture patient improvement. We defined ASAS50 and ASAS70 as a 50% and 70% relative improvement and an absolute improvement ≥20 units in 3 of 4 domains with no worsening in the fourth domain. It is possible that if the absolute level of improvement was set higher, then agreement between the expert definitions of improvement and ASAS would have been better than that observed.

Expert criteria for improvement based on the physician’s global assessment of disease activity did not agree completely with the ASAS criteria. On the other hand, there was good agreement between expert criteria for improvement based on patient global scores and ASAS. This may result from the fact that patient global assessment is part of the 4 variables that define ASAS improvement. Furthermore, little agreement was found between expert definition of improvement based on physician global scores and that based on patient global scores, potentially because there is little agreement between physicians and patients’ conceptual approach to disease activity. We believe that the patient is ultimately the best judge of the degree of disease activity and that, therefore, the ASAS criteria are a valid set of criteria for the assessment of response to biologic agents in AS patients.

A potential criticism of this study is that we compared expert clinicians’ definition of improvement based on patient global assessment of disease activity to ASAS. The concern with this approach, as mentioned above, is that a physician’s perception of a dramatic improvement may differ from that of a patient’s. However when we sought a definition of a dramatic response from a sample of 15 AS patients treated with biologic agents, we learned that the AS patients’ perception of a dramatic improvement was exactly comparable to the expert definition of dramatic improvement based on patient global assessment of disease activity scores.

In this study we have shown that the ASAS relative

<table>
<thead>
<tr>
<th>Table 1. Stratification of magnitude of clinical response as defined by expert consensus based on changes in patient global and physician global assessment scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome variable</strong></td>
</tr>
<tr>
<td>Patient global assessment of disease activity*</td>
</tr>
<tr>
<td>Physician global assessment of disease activity†</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* Five-point Likert scale anchored at 0 = no disease activity to 4 = very severe disease activity.
† Four-point Likert scale anchored at 0 = no disease activity to 3 = severe disease activity.

| Table 2. Two-by-two tables comparing expert definition of dramatic and good improvement based on changes in ratings of physician and patient global assessment scores to ASAS criteria of improvement* |
|-------------------|-------------------|-------------------|
| **Dramatic improvement** | **ASAS70** | **No** |
| Expert criteria based on physician global scores | Yes | 4 | 4 | $\kappa = 0.21$ |
| | No | 8 | 24 | $\kappa = 0.21$ |
| Expert criteria based on patient global scores | Yes | 5 | 0 | $\kappa = 0.60$ |
| | No | 7 | 28 | $\kappa = 0.60$ |
| **Good improvement** | **ASAS50** | **No** |
| Expert criteria based on physician global scores | Yes | 15 | 6 |
| | No | 7 | 12 | $\kappa = 0.35$ |
| Expert criteria based on patient global scores | Yes | 17 | 1 |
| | No | 5 | 17 | $\kappa = 0.70$ |

* See Table 1 for expert definitions of improvement based on physician and patient global assessment of disease activity. ASAS = Ankylosing Spondylitis Assessment Study Group.
† ASAS70 was defined as a 70% relative improvement and at least 10 units of absolute improvement in 3 of 4 of the ASAS core domains with no worsening in the fourth domain and was considered, by expert consensus, to constitute a dramatic improvement for the purpose of this exercise.
‡ ASAS50 was defined as a 50% relative improvement in and at least 10 units of absolute improvement 3 of 4 of the ASAS core domains with no worsening in the fourth domain and was considered, by expert consensus, to constitute a dramatic improvement for the purpose of this exercise.
improvement criteria are valid criteria to use in the detection of improvements experienced by AS patients treated with TNF-α-blocking agents. On the other hand, a definition of improvement based on changes in patients’ global assessment of disease activity may also be adequate, particularly for use in the routine clinical setting. The patient’s global assessment of disease activity represents a simple and inexpensive outcome measure that reflects improvement noted in patients with AS taking biologic agents. However, further study is warranted to validate these claims against an objective measure of improvement, such as erythrocyte sedimentation rate, C-reactive protein, or appropriate imaging.

REFERENCES