Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis

Margaret C. Reilly¹, Katherine L. Gooch², Robert L. Wong³, Hartmut Kupper⁴ and Désirée van der Heijde⁵

Abstract

Objective. To determine the validity, reliability and responsiveness of the Work Productivity and Activity Impairment questionnaire in AS (WPAI:SpA).

Methods. Baseline and Week-24 data from a randomized, double-blind study of adalimumab in patients with AS were used. Discriminative validity of WPAI:SpA absenteeism, presenteeism, overall work productivity loss and activity impairment scores was assessed relative to patient-reported outcomes: Bath AS Disease Activity Index (BASDAI), AS Quality of Life Questionnaire (ASQOL), Short-Form 36 Health Survey (SF-36), Physical and Mental Component Summaries (PCS and MCS, respectively) and Health Utilities Index Mark 3 (HUI-3). Responsiveness of the WPAI:SpA instrument was assessed for patients meeting the minimum clinically important differences (MCIDs) for ASQOL and BASDAI (i.e. quality of life and clinical responders, respectively) and quantified with standardized response mean (SRM) calculations.

Results. Of 315 patients, 205 were employed at baseline. Patients with more severe AS (BASDAI > median) showed significantly greater impairment in work and daily activities than patients with lesser disease severity (P < 0.001). This trend was consistent for ASQOL, SF-36 PCS, SF-36 MCS and HUI-3. There were significant differences in WPAI:SpA scores for patients achieving BASDAI clinical response and ASQOL quality of life response vs non-responders. For responders, SRMs were large for work presenteeism, overall work impairment and activity impairment (−0.86 to −1.29 for BASDAI; −0.89 to −1.18 for ASQOL) and small for absenteeism (−0.25 for BASDAI; −0.31 for ASQOL).

Conclusions. The WPAI:SpA is a valid, reliable and responsive tool for assessing work productivity for patients with AS.


Key words: Adalimumab, AS, Work productivity, WPAI, Patient-reported outcome measures.

Introduction

AS is a chronic, rheumatic inflammatory disease of the axial skeleton, large peripheral joints and entheses. The average age at onset of AS is between 20 and 40 years [1–3]. Men are affected by AS three times more often than women [4]. Disease progression results in pain, joint stiffness and varying degrees of spinal fusion, causing considerable functional limitations [5]. Owing to the early onset and long-term functional disability caused by AS, there is a significant impact on health-related quality of life (HRQOL) [5]. In addition, over the course of disease, withdrawal from the work force has been estimated as being three times greater in patients with AS than in the general population [6] and decreased productivity at work has been reported in patients with AS [7]. The functional disability and decreased work productivity observed in patients with AS are associated with substantial
psychosocial and financial consequences for the affected patients as well as society in general [8–11].

Although decreased work productivity in patients with AS has been demonstrated in several studies, an analysis of several of these studies has concluded that there is considerable variability in the findings, a result, in part, of the large variation in endpoint definitions [4]. A standardized measurement of work productivity, validated in patients with AS, would allow consistent assessment of the impact of disease on work productivity across different AS subgroups, countries and cultures.

The Work Productivity and Activity Impairment (WPAI) questionnaire is a self-administered instrument used to assess the impact of disease on productivity [12]. The WPAI measures work productivity loss due to general health or a specified health problem [13]. This questionnaire has been validated for use in many diseases, including allergic rhinitis [14], dermatitis [15], gastro-oesophageal reflux disease [16], nocturia [17], asthma [18], irritable bowel syndrome [19] and Crohn’s disease [20].

The purpose of this analysis was to determine whether the WPAI modified for AS (WPAI:SpA) is a valid and responsive instrument for assessing work productivity in patients with AS based on psychometric analysis. We have provided definitions of the key psychometric terms used in this paper to assist clinicians with the understanding of their usage in the text to follow. Construct validity is the validity of a test or a measurement tool established by demonstrating its ability to identify or measure the variables or constructs that it proposes to identify or measure. The judgement is frequently based on the degree of correlation between the instrument being evaluated and more established measurements of the same construct. In other words, construct validity is often based on assessments of agreement between an instrument and an acceptable or ‘gold standard’ measurement of the health state of interest. Responsiveness is defined as a tool’s sensitivity to differences in clinical changes over time, whereas reliability is the degree of stability of the tool when there is no clinical change. In this context, responsiveness does not refer to the impact a particular treatment or therapy has on a disease state.

**Methods**

**Patient enrolment and study design**

Patients for this study were enrolled in a placebo-controlled, double-blind, randomized, multicentre study of adalimumab vs placebo conducted in the USA and Europe at 43 sites [Adalimumab Trial Evaluating Long-term Efficacy and Safety in AS (ATLAS)] (ClinicalTrials.gov Identifier: NCT00085644) [21–23]. Eligible patients were at least 18 years of age with a diagnosis of AS according to the modified New York criteria [24] and had an inadequate response or intolerance to one or more NSAIDs and in addition may have failed therapy with one or more DMARDs [21–23]. Patients also met at least two of the following three criteria: a Bath AS Disease Activity Index (BASDAI) score of ≥4, a total back pain score of ≥4 [visual analogue scale (VAS) of 0–10 cm], or a duration of morning stiffness ≥1 h. Patients were randomly assigned in a 2:1 ratio to receive either 40 mg adalimumab every other week subcutaneously or matching placebo during a 24-week, placebo-controlled, double-blind period.

Informed consent was obtained from each patient. Each participating site received ethics approval, and the study was conducted according to the International Conference on Harmonization guidelines and the Declaration of Helsinki. A full description of the methods utilized in ATLAS is reported elsewhere [21–23].

**Patient-reported outcome measures**

Patients completed self-administered questionnaires at baseline and at defined subsequent time points throughout the ATLAS study. The self-administered questionnaires included the WPAI:SpA, the BASDAI, the AS Quality of Life Questionnaire (ASQOL), the Short-Form 36 Health Survey (SF-36) and the Health Utilities Index Mark 3 (HUI-3). Only baseline and Week-24 questionnaire data are reported here.

The WPAI is in the public domain and can be used without fee or permission from the authors [25]. The WPAI is constructed such that it can be modified for any health problem by specifying the disease/condition of interest in the questions. For the WPAI:SpA employed in this study, we specified AS in the WPAI template. The WPAI:SpA assesses the impact of AS on work and other daily activities during the past 7 days [12]. The WPAI:SpA consists of six questions to determine employment status, hours missed from work due to AS, hours missed from work for other reasons, hours actually worked, the degree to which AS affected work productivity while at work and the degree to which AS affected activities outside of work. Four scores are derived: percentage of absenteeism, percentage of presenteeism (reduced productivity while at work), an overall work impairment score that combines absenteeism and presenteeism and percentage of impairment in activities performed outside of work. Greater scores indicate greater impairment. Questions related to absenteeism and presenteeism were applicable to employed patients only. The WPAI:SpA employed in this study is provided in supplementary Appendix S1, available as supplementary data at *Rheumatology* Online. In addition, scoring instructions for the WPAI:SpA are provided in supplementary Appendix S2.

The BASDAI is a simple, six-item questionnaire designed to assess the disease activity level of AS during the past week [26]. The BASDAI consists of six questions rated on a 10-cm horizontal VAS that measures severity of the five major symptoms of AS: fatigue, spinal joint pain, peripheral joint pain, localized tenderness and morning stiffness (duration and severity) [26]. The BASDAI is calculated by adding the VAS results for each symptom, using the mean of the two morning stiffness scores. The total score is divided by 5 and the BASDAI is reported on a
The ASQOL is designed to assess the impact of AS on the patient’s quality of life [27, 28]. The ASQOL consists of 18 questions (‘yes’ or ‘no’ response) related to the impact of AS pain on patients’ ability to cope, relationships, mood, sleep, motivation, activities of everyday living, independence and social life at that point in time. The ASQOL is reported on a scale of 0–18, with lower ASQOL scores indicating a better quality of life [27, 28].

The SF-36 is a generic measure of health status outcomes used in general and disease-specific populations [29, 30]. The SF-36 version applied in ATLAS asked patients to recall their experiences in the past 4 weeks. The SF-36 standardized scoring system yields eight health scores and two (physical and mental) composite summary scores: the Physical Component Summary (PCS) and Mental Component Summary (MCS) [29, 30]. Greater SF-36 scores indicate better quality of life. The SF-36 PCS and MCS scores were utilized for the present study.

The HUI-3 is a preference-based, generic health status classification system that incorporates both qualitative and quantitative aspects of health. Patients are asked to recall information regarding specific health aspects in the past 4 weeks. The HUI-3 consists of 16 questions that assess HROQL in the following eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain, with five or six levels per attribute [31–33]. A greater HUI score indicates better quality of life.

Statistical analyses
The known group validation method was used to assess the construct validity (see definition at the end of the ‘Introduction’ section) of the WPAI:SpA [20]. Groups of patients were constructed using the baseline median scores of the BASDAI, ASQOL, SF-36 PCS and SF-36 MCS, and the overall HUI-3 score. Patients were expected a priori to have less impairment in work and activities if they (i) were experiencing fewer AS signs and symptoms (BASDAI < median), (ii) had reported better AS-specific quality of life (ASQOL < median) and (iii) demonstrated better generic quality of life (SF-36 PCS > median, SF-36 MCS > median and HUI-3 > median).

The medians were calculated for employed patients to determine the validity of the absenteeism, presenteeism and overall work impairment scores. The medians were calculated for all patients to determine the validity of the daily activity impairment score. Two-sided non-parametric Wilcoxon rank tests were applied to compare the patients above the median with those below the median. Only baseline data were utilized for the validation analysis.

WPAI:SpA responsiveness is, as noted, defined as the instrument’s sensitivity to differences in health status changes over time [34], rather than responsiveness of a patient with AS to a given therapy. The responsiveness of the WPAI:SpA was calculated using baseline and Week-24 data. Responsiveness criteria were determined using established clinically meaningful improvements in the ASQOL and BASDAI, which is similar to the approach used to establish the responsiveness of WPAI measures for Crohn’s disease [20]. The minimum clinically important difference (MCID) of the ASQOL has been identified as a change of −1.8 on the 18-point ASQOL scale [35] (i.e. patients meeting this criterion are ‘quality of life responders’), and the MCID for the BASDAI has been determined as a change of −1.96 on the 10-point BASDAI scale (i.e. patients meeting this criterion are ‘clinical responders’) [36]. The changes in WPAI:SpA scores at baseline and Week 24 were calculated according to the ASQOL and BASDAI response criteria and compared using two-sided Wilcoxon tests. Responsiveness was quantified by calculating the standardized response mean (SRM) in each group. This measure of effect size is judged as small when <0.5, moderate when between 0.5 and 0.8 and large when >0.8 [37].

The instrument reliability (see definition at the end of the ‘Introduction’ section) of the WPAI:SpA was tested within the context of the responsiveness analysis by comparing the change in scores of patients not achieving MCID (i.e. non-responders) with the change in scores of patients achieving MCID (i.e. responders).

Results
Patient characteristics at baseline
A total of 315 patients were randomly selected and enrolled in the study. The mean age was 42.2 years, 74.9% were males and the average duration of AS was 10.9 years (Table 1). The mean baseline BASDAI score for these 315 patients was 6.3 ± 1.7 and the mean baseline ASQOL score was 10.3 ± 4.3 (Table 2). The baseline characteristics of all patients were similar compared with only those patients employed at baseline (n = 205), apart from a greater distribution of male patients in the employed group. Tests for statistical differences were not undertaken because employed patients were included in both the groups.

<table>
<thead>
<tr>
<th>Table 1 Demographics and clinical characteristics of all patients and employed patients at baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients (n = 315)</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Age, mean (range), years</td>
</tr>
<tr>
<td>Male, n (%)</td>
</tr>
<tr>
<td>White, n (%)</td>
</tr>
<tr>
<td>Duration of AS, mean (s.d.), years</td>
</tr>
</tbody>
</table>
Baseline productivity and patient-reported outcomes

At baseline, 65.1% of the patients were employed. For these patients, the mean BASDAI was 6.1 ± 1.7 and the mean ASQOL was 9.4 ± 4.1. The amount of missing or invalid data at baseline was minimal; WPAI:SpA scores were calculated for >88% of patients (range across subscales 88.4–97.8). Mean absenteeism was 9.0%, mean presenteeism was 41.7% and mean overall work impairment was 43.9% (Table 2). A total of 26.3% of employed patients reported absenteeism during the previous 7 days and 87.8% reported presenteeism. Of all patients, 97.1% reported some impairment in non-work activities during the past 7 days.

Constructive validity

Median scores for the ASQOL, BASDAI, SF-36 PCS and MCS, and HUI-3 were used to discriminate between patients with the worst health (worst AS disease severity) and those with the best health (least AS disease severity) at baseline (Table 3). Patients with AS of the worst severity...
Based on two-sided Wilcoxon test comparing mean change from baseline to Week 24.

At baseline, patients showed significant clinical signs of AS disease activity, defined as a 1.96-point decrease in BASDAI score. BASDAI clinical responders had a significant reduction in work impairment and daily activity impairment (presenteeism/P<0.001 and overall work impairment/P<0.001), as assessed by the WPAI:SpA, than patients with lesser disease severity. Absenteeism was slightly higher, although not significantly so, for patients with the worst health compared with patients with the best health (difference−0.4, P=0.35).

Patients with the best health, as measured by ASQOL (ASQOL<median), had significantly lower WPAI:SpA impairment scores: absenteeism (difference−12.0, P<0.001), presenteeism (difference−22.6, P<0.001), overall work impairment (difference−21.2, P<0.001) and daily activity impairment (difference−26.3, P<0.001). Similarly, the patients with the worst health, as defined by the SF-36 PCS<median, SF-36 MCS<median and HUI-3<median, had statistically significant greater work productivity loss and activity impairment than the patients with the best health for absenteeism, presenteeism, overall work impairment and daily activity impairment (P=0.002 to <0.001), except for absenteeism comparing worst with best health with the SF-36 MCS (P=0.08).

Responsiveness and reliability

Change in WPAI:SpA from baseline to Week 24 by BASDAI clinical response. Of the 315 patients enrolled in ATLAS, 299 patients (including 197 employed patients) with complete baseline and Week-24 data were included in the present study. A total of 197 (65.89%) of all 299 ATLAS patients and 141 (71.57%) of 197 employed ATLAS patients showed a clinically significant improvement at Week 24 in AS disease activity, defined as a 1.96-point decrease in BASDAI score. BASDAI clinical responders had a significant reduction in work impairment and daily activity impairment (presenteeism−20.3, P=0.04; overall work impairment−21.8, P=0.05; daily activity impairment−28.0, P<0.001) compared with patients with AS who did not achieve BASDAI clinical response (Table 4). The SRMs for the patients with improvement in disease activity in these WPAI:SpA scores were large (−0.86 to −1.29). Although the reduction in absenteeism was greater in patients with improved disease activity compared with patients with no improvement, this change was not significant (difference−1.8, P=0.11). BASDAI clinical non-responders demonstrated small improvements in WPAI:SpA scores, but the SRMs were small or moderate (−0.14 to −0.54) (Table 4).

Change in WPAI:SpA from baseline to Week 24, by ASQOL quality of life response. A total of 212 (70.90%) of all 299 ATLAS patients with complete baseline and Week-24 data and 141 (71.57%) of 197 employed ATLAS patients with complete baseline and Week-24 data showed a clinically significant improvement at Week 24 in AS disease activity, defined as a 1.8-point improvement in ASQOL score. ASQOL quality of life responders had a significant reduction in work impairment and daily activity impairment (presenteeism−5.9, P=0.04; overall work impairment−23.4, P=0.02; daily activity impairment−26.6, P<0.001), compared with patients who did not achieve ASQOL response (Table 5). For patients with an improvement in disease-specific quality of life, the SRM was small (−0.31) for absenteeism and large for the other WPAI:SpA scores (−0.89 to −1.18). ASQOL non-responders had small changes in their WPAI:SpA scores (−0.11 to −0.46) (Table 5).

Discussion

The adverse impact of AS on socio-economic factors such as work productivity is a significant burden on patients with AS. For determining whether treatment can ameliorate this burden, a validated tool that measures work productivity is necessary. For this reason, we tested the validity, reliability and responsiveness of the WPAI:SpA in patients participating in a clinical trial of AS.

At baseline, patients showed significant clinical signs and symptoms of AS, as measured by the disease-specific BASDAI and ASQOL. In addition, the SF-36 MCS and SF-36 PCS scores were low compared with the general US population [38], indicating decreased overall physical and mental health status in patients with AS. We found that patients with worse disease severity...
Table 5 Mean (s.d.) change in WPAI:SpA questionnaire from baseline to Week 24, by ASQOL response

<table>
<thead>
<tr>
<th>WPAI:SpA measurement</th>
<th>ASQOL responsea &lt;1.8</th>
<th>ASQOL responsea &gt;1.8</th>
<th>Difference, ASQOL quality of life responders vs non-responders</th>
<th>$P$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absenteeismb</td>
<td>42</td>
<td>1.6 (14.3)</td>
<td>124</td>
<td>5.9 (19.3)</td>
</tr>
<tr>
<td>Presenteeismb</td>
<td>43</td>
<td>−10.7 (23.5)</td>
<td>114</td>
<td>−21.1 (23.7)</td>
</tr>
<tr>
<td>Overall work impairmentb</td>
<td>40</td>
<td>−9.6 (25.6)</td>
<td>116</td>
<td>−23.4 (24.9)</td>
</tr>
<tr>
<td>Daily activity impairmentb, %</td>
<td>82</td>
<td>−7.6 (18.9)</td>
<td>206</td>
<td>−26.6 (22.5)</td>
</tr>
</tbody>
</table>

$^a$MCID for ASQOL clinical response, 1.8-point decrease on the 18-point scale. $^b$Employed patients. $^c$All patients. $^*P$-value based on two-sided Wilcoxon test comparing mean change from baseline to Week 24.

and lesser HRQOL scores also had significantly more absenteeism, presenteeism, overall work productivity loss and activity impairment than patients with less severe disease and greater HRQOL scores. These findings are consistent with the validation of WPAI in other specific diseases, such as Crohn’s disease, irritable bowel syndrome, allergic rhinitis, dermatitis and gastro-oesophageal reflux disease [15, 16, 19, 20, 39].

We also sought to determine whether the WPAI:SpA was responsive by evaluating the change in WPAI:SpA results from baseline to Week 24, as related to improvements in measurements of disease activity and HRQOL that have been previously established as clinically meaningful. The changes in WPAI:SpA scores from baseline to Week 24 in ASQOL quality of life responders and BASDAI clinical responders were superior to changes in WPAI:SpA scores in patients not meeting the respective ASQOL or BASDAI response criteria. In addition, the effect sizes for presenteeism, overall work impairment and overall daily activity impairment for quality of life non-responders were small, whereas they were large for quality of life responders. The effect size for absenteeism in quality of life responders, although small, was larger than in quality of life non-responders ($P = 0.04$ for effect size according to ASQOL response; $P = 0.11$ according to BASDAI response); this finding is similar to findings reported in other WPAI validation studies [15, 19, 20]. A potential reason for this small effect size for absenteeism is that patients with a chronic condition such as AS may have difficulty taking time off from work, which may result in increased presenteeism, as our results suggest. Alternatively, because the average duration of AS at baseline in this study was 10.6 years (employed patients), the low absenteeism may be a result of these patients finding and adapting to jobs that are compatible with the disease manifestations of AS.

The results of this study demonstrated that the WPAI:SpA is a valid, reliable and responsive instrument to measure the impact of AS on work absenteeism and overall productivity. The results conform with the OMERACT filter, whereby the WPAI:SpA demonstrates truth (i.e. construct validity), discrimination (i.e. able to discriminate between groups and is responsive to change) and feasibility (i.e. is easily applied as demonstrated by minimal missing or invalid data) [40]. Assuming a 40 h work week, the 13.8% lower overall work impairment for those meeting the ASQOL quality of life response criterion of $−1.8$ is equivalent to 5.5 h of increased productivity per week relative to ASQOL non-responders. Similarly, 6.5% lower overall work impairment for those meeting the BASDAI clinical response criterion of $−1.96$ is equivalent to a 2.6 h per week improvement in productivity compared with BASDAI clinical non-responders. Overall, the results of this study suggest that the WPAI is a relevant and appropriate tool for clinicians to consider in assessing the burden of AS on employed patients. The responsiveness analysis also suggests that work productivity is directly related to disease activity.

There are limitations of this study. First, social system structures observed in different countries may influence the validation of the WPAI:SpA for a given country. In this study, each country sample size was not powered to assess the differences in results. Secondly, the ATLAS study limited inclusion of patients with AS to those with at least two of the following disease severity criteria: BASDAI $\geq 4$, total back pain VAS score $\geq 4$ and duration of morning stiffness $\geq 1$ h. Therefore, it is unknown whether these results can be generalized to patients with less severe AS disease. However, 11.7% of ATLAS patients (37 of the 315 patients enrolled) had baseline BASDAI scores $<4$, suggesting that these results might apply to patients with less severe disease. In addition, the patient-reported hours of absenteeism or presenteeism could not be validated against employment records. Studies to determine the correlation between patient-reported and more objective measurements of work productivity are very difficult to execute and go beyond the scope of activity in drug efficacy and safety trials [20].

Finally, this analysis was not designed to evaluate the impact adalimumab had on improving work productivity, as observed in the ATLAS study. Such results have been published [41], along with analyses of which patient characteristics and other factors were found to be associated with work productivity. This analysis determined that younger age and male sex were significantly and...
independently associated with working patients with AS in ATLAS [41]. In addition, SF-36 PCS, ASQOL and HUI-3 scores, and both patient’s global assessment of disease activity and nocturnal pain scores were independently associated with working status. Work absenteeism was weakly correlated with patient-reported outcome scores. Moreover, WPAI components of work presenteeism (lack of productivity at work), activity impairment and overall work productivity loss due to AS were moderately correlated with HRQOL, as measured by the ASQOL, SF-36 PCS and SF-36 Bodily Pain domain. Linear multivariate analyses showed that presenteeism was significantly associated with pain, functioning and disease activity. In this study, long-term adalimumab therapy (through 3 years of exposure) was associated with sustained improvements in WPAI [41]. In summary, the discriminative validity of the WPAI:SpA as well as its responsiveness to clinically meaningful change was established, making the WPAI:SpA the only validated tool for assessing work productivity loss in patients with AS.

Rheumatology key messages

- WPAI:SpA is a valid and responsive tool for assessing work productivity in patients with AS.
- Patients with greater disease severity show greater work impairment.

Acknowledgements

The authors thank all members of the ATLAS Study Group for their contributions to this long-term study and their continued care of all participating patients and Abbott personnel for organizing the study. The authors thank Hoa Khong, MD, of VComTech, Howard R. Hubbell, PhD, of Hubbell Consulting, LLC and Michael A. Nissen, ELS, of Abbott Laboratories, for assistance with analysis, manuscript preparation and revision.

Funding: This work was supported by Abbott Laboratories, Abbott Park, IL, USA.

Disclosure statement: M.C.R. is a consultant for Abbott Laboratories, Amgen, AstraZeneca, Medimmune, UCB and Wyeth. K.L.G. and H.K. (holds shares and stock options) are employees of Abbott Laboratories. R.L.W. was an Abbott employee at the time this analysis was conducted. D.vdH. has received consulting fees from Abbott Laboratories, Amgen, Aventis, Bristol Meyers Squibb, Centocor, Pfizer, Roche, Schering Plough, UCB and Wyeth.

Supplementary data

Supplementary data are available at Rheumatology Online.

References


