

Association of Improvement in Pain With Therapeutic Response as Determined by Individual Improvement Criteria in Patients With Rheumatoid Arthritis

EVA C. SCHARBATKE,¹ FRANK BEHRENS,² MARC SCHMALZING,¹ MICHAELA KOEHM,² GERD GREGER,³ HOLGER GNANN,⁴ HARALD BURKHARDT,² AND HANS-PETER TONY¹

Objective. To use statistical methods to establish a threshold for individual response in patient-reported outcomes (PROs) in patients with rheumatoid arthritis.

Methods. We used an analysis of variance model in patients on stable therapy (discovery cohort) to establish critical differences (d_{crit}) for the minimum change associated with a significant individual patient response (beyond normal variation) in the PRO measures of pain (0–10), fatigue (0–10), and function (Funktionsfragebogen Hannover questionnaire; 0–100). We then evaluated PRO responses in patients initiating adalimumab in a noninterventional study (treatment cohort).

Results. In the discovery cohort ($n = 700$), PROs showed excellent long-term retest reliability. The minimum change that exceeded random fluctuation was conservatively determined to be 3 points for pain, 4 points for fatigue, and 16 points for function. In the treatment cohort ($n = 2,788$), 1,483 patients (53.2%) achieved a significant individual therapeutic response as assessed by Disease Activity Score in 28 joints (DAS28)– d_{crit} (≥ 1.8 points) after 12 months of adalimumab treatment; 68.5% of patients with a DAS28– d_{crit} response achieved a significant improvement in pain, whereas approximately 40% achieved significant improvements in fatigue or function. Significant improvements in all 3 PROs occurred in 22.7% of patients; 22.8% did not have any significant PRO responses. In contrast, significant improvements in all 3 PROs occurred in only 4.4% of 1,305 patients who did not achieve a DAS28– d_{crit} response at month 12, and 59.1% did not achieve any significant PRO responses.

Conclusion. The establishment of critical differences in PROs distinguishes true responses from random variation and provides insights into appropriate patient management.

INTRODUCTION

Patient-reported outcomes (PROs) provide an important perspective on the patient's perceptions of his or her clinical well-being. These outcomes, including patient global

disease assessment, pain, function, and fatigue can be used to accurately discriminate between active treatment and placebo in patients with rheumatoid arthritis (RA) and may be more sensitive to treatment effectiveness than objective clinical assessments (1). Furthermore, certain

ClinicalTrials.gov identifiers: NCT01078090, NCT01077258.

Supported by AbbVie Deutschland GmbH & Co. KG (formerly Abbott GmbH & Co. KG).

¹Eva C. Scharbatke, MD, Marc Schmalzing, MD, Hans-Peter Tony, MD: Schwerpunkt Rheumatologie/Klinische Immunologie Medizinische Klinik und Poliklinik II, Universität Würzburg, Würzburg, Germany; ²Frank Behrens, MD, Michaela Koehm, MD, Harald Burkhardt, MD: University Hospital Frankfurt, Goethe University, and Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine & Pharmacology TMP, Frankfurt am Main, Germany; ³Gerd Greger, PhD: AbbVie Deutschland GmbH & Co. KG, Wiesbaden, Germany; ⁴Holger Gnann,

Dipl-Stat: Abteilung Biostatistik, GKM Gesellschaft für Therapieforchung mbH, München, Germany.

Drs. Scharbatke, Behrens, Burkhardt, and Tony contributed equally to this work.

Dr. Scharbatke has received speaker's fees from AbbVie Deutschland GmbH & Co. KG (less than \$10,000). Dr. Behrens has received speaking fees, research funding, and consultant fees from AbbVie Deutschland GmbH & Co. KG (less than \$10,000). Dr. Schmalzing has received speaking fees and consultant fees from AbbVie Deutschland GmbH & Co. KG (less than \$10,000). Dr. Koehm has received speaking fees, research funding, and consultant fees from AbbVie Deutschland GmbH & Co. KG (less than \$10,000).

Significance & Innovations

- We used statistical methods to determine a critical difference (d_{crit}) required for significant individual patient responses in patient-reported outcomes (PROs) in patients initiating therapy with adalimumab, a tumor necrosis factor inhibitor, and to evaluate correlations between PROs (pain, fatigue, and function) and objective measures of disease activity, such as the Disease Activity Score in 28 joints (DAS28).
- Of the 3 PROs evaluated, pain showed the highest correlation with DAS28 and fatigue showed the lowest. Of patients with an objective therapeutic response as defined by DAS28- d_{crit} at month 12, 68.5% showed a significant individual improvement in pain and approximately 40% showed improvements in fatigue and function.
- Compared with patients who had an objective DAS28 response to therapy, patients who did not experience an objective DAS28 response to therapy were far less likely to achieve significant improvements in PROs.
- Our data supply additional support for the relevance of statistically determined individual improvement criteria in clinical practice and provide insights into appropriate patient management and typical outcomes during adalimumab therapy that may be useful in managing treatment expectations.

PROs, including patient global disease assessment and functional questionnaires, are more closely related to clinical outcomes than objective assessments of disease activity such as laboratory markers, joint counts, or radiographs (2). In multiple studies, patient-reported functional status has been shown to be the most significant predictor of mortality in patients with RA, outweighing clinical or laboratory assessments such as joint counts, rheumatoid factor seropositivity, or erythrocyte sedimentation rate (ESR) (reviewed by Pincus et al [2]).

Because of the importance of PROs in evaluating treatment response and predicting long-term outcomes, we utilized data from a large German noninterventional study of patients with RA receiving adalimumab during routine clinical care (3) to examine the association between PROs

and the Disease Activity Score in 28 joints (DAS28), a frequently used measure of disease activity that includes 3 objective components (swollen joint count, tender joint count, and ESR) and 1 subjective component (patient global assessment of health or disease activity [PGA]). In a previous publication, we reported on the use of statistical measures to determine a critical difference (d_{crit}) for the minimum change associated with a significant individual patient response in disease activity as assessed by the DAS28 (4). This method is fundamentally different from those used to evaluate minimum clinically important differences (MCIDs). The goal of MCID assessments is to identify the smallest change perceived as beneficial by patients (5), whereas the d_{crit} method is based on the premise that significant individual responses should, at minimum, result in changes that exceed within-subject, measurement-associated variability over periods of time relevant to a clinical setting (4). In this respect, the d_{crit} method is similar to assessments of the smallest detectable difference (SDD) (6) or the reliable change index (7). However, these latter assessments are usually based on reliability constants derived from measurements made over a short period of time, rather than over a period of months to a year, a time frame that is more relevant to assessment of RA.

Here we utilize the d_{crit} statistical method to establish criteria for significant responses in the PROs of pain, fatigue, and function. Our study provides key insights into the magnitude of change required for a significant PRO response and into the association between therapeutic response and significant improvements in PROs.

PATIENTS AND METHODS

Patients. This study utilized 2 German patient cohorts: a discovery cohort including only patients with stable disease and on stable therapy, and a treatment cohort. Both patient cohorts were derived from multicenter, noninterventional studies of patients with active RA initiating treatment with adalimumab, a tumor necrosis factor inhibitor (TNFi).

Patients in the discovery cohort were enrolled in a 2-year noninterventional study (NCT01077258), which has been previously described (3), and were seen between April 2004 and February 2013. Only patients with stable therapy (no changes in therapeutic agents or corticosteroid dose) between month 12 and month 24 of the study were included in discovery cohort analyses. In addition, discovery cohort patients were required to have a complete data set (data at months 12, 18, and 24) for 1 or more PRO variable. For each PRO evaluation, analyses were limited to patients in the discovery cohort with a complete data set for the PRO being evaluated. Data from months 12, 18, and 24 were used to assess long-term test-retest reliability of PRO measures.

Patients in the treatment cohort were enrolled in a 5-year noninterventional study (NCT01078090) with a similar design to the 2-year study. These patients were seen between April 2003 and March 2013. To be included in the noninterventional studies, patients were required to have a diagnosis of RA, active disease, a clinical

Dr. Burkhardt has received speaking fees, research funding, and consultant fees from AbbVie Deutschland GmbH & Co. KG (less than \$10,000). Dr. Tony has received speaking and consulting fees and compensation for board memberships from AbbVie Deutschland GmbH & Co. KG (less than \$10,000).

Address correspondence to Hans-Peter Tony, MD, Medizinische Klinik und Poliklinik II Universitätsklinikum Würzburg, Rheumatologie/Klinische Immunologie, Oberdurrbacher Strasse 6, 97080 Würzburg, Germany. E-mail: Tony_H@ukw.de.

Submitted for publication August 10, 2015; accepted in revised form March 8, 2016.

indication for treatment with a TNFi, and no contraindications to TNFi therapy. All patients were informed of the study objectives and gave written consent for the anonymous use of their personal data in statistical analyses. Because of the noninterventional nature of this study, ethics approval was not required by German law.

Patients in the treatment cohort who were included in the analyses reported here were required to have data for DAS28, pain, function, and fatigue at baseline and month 12. Patients receiving additional biologic therapies and those who had received previous therapy with adalimumab were excluded from these analyses.

Evaluations of disease activity and PROs. Disease activity was assessed by the DAS28, a validated instrument in which higher scores indicate greater disease activity (8,9). The self-administered Funktionsfragebogen Hannover (FFbH) patient questionnaire was used to assess patient function on a scale of 0 (total loss of functional capacity) to 100 (maximal functional capacity) units (10); the FFbH score indicates the remaining percentage of function. The FFbH has been validated in patients with RA and is comparable to the Health Assessment Questionnaire Disability Index (11); an English translation has been published by Westhoff et al (12). Pain and fatigue were evaluated by use of 11-point scales in which a score of 0 represented the best possible status and a score of 10 indicated the worst possible status. PGA was not evaluated as a PRO, as our goal was to examine the association between PROs and DAS28. Because PGA is one of the measures used to calculate DAS28, the use of this PRO would have confounded the analyses. However, PGA and physician global assessment of disease activity (MDGA) were included in evaluations of correlations between PROs and other measures of disease activity. PGA and MDGA were assessed on 11-point scales, in which a score of 0 represented the best possible status and a score of 10 indicated the worst possible status.

Statistical analyses. Summary statistics are presented for demographic and disease characteristics. Missing data were not imputed. The method for determining the long-term reliability and critical difference (e.g., the minimum change that can be reliably discriminated from random variations) in PROs was based on evaluations of intraindividual variation in patients who were treated for 2 years, with stable therapy from months 12 through 24, as has been described previously (4). Briefly, we adapted the method of Lienert and Raatz (13) to determine a critical difference based on the 1-sided 5% z-value of the normal distribution in patients on stable therapy. The 1-sided critical difference (i.e., d_{crit} value) was used to define therapeutic response (14), as only reductions in disease activity are relevant to defining a response. D_{crit} values for increases in disease activity were not evaluated. The 1-sided d_{crit} values for reductions in disease activity were then used to evaluate responses in patients initiating adalimumab therapy. Patients who showed an improvement that equaled or exceeded the 1-sided d_{crit} value were considered to have experienced a statistically significant improvement that could not be explained by intraindividual

variation and were classified as responders for that outcome. Patients with changes less than the d_{crit} value or with increases in disease activity were classified as nonresponders for that outcome. The method of Lienert and Ratz (13) is based on within-subject variance rather than between-subject variance.

Pearson's correlation coefficients were used to evaluate the association between PROs and other disease activity assessments (1 indicates a perfect positive linear relationship and -1 indicates a perfect negative linear relationship).

RESULTS

Determination of the critical difference in PROs. The discovery cohort ($n = 700$) consisted of patients who had received stable RA therapy (no changes in therapeutic agents or corticosteroid dose) between months 12 and 24 after initiating adalimumab therapy in a noninterventional study. This requirement for stable treatment allowed intraindividual fluctuations in outcomes to be distinguished from responses due to alterations in therapy. Demographic and disease characteristics of the patients in the discovery cohort are presented in Table 1 for both baseline (initiation of adalimumab treatment) and month 12 (entry into the discovery cohort on the basis of stable disease and therapy). Patients had a mean disease duration of 11.4 years and a mean age of 54.3 years at baseline. After 12 months of adalimumab treatment, patients in the discovery cohort had low to moderate disease activity (mean DAS28 score 3.3). The discovery cohort maintained a mean \pm SD DAS28 of 3.3 ± 1.4 at month 18 and month 24, thereby achieving the crucial methodologic requirement of constant disease activity under fixed therapeutic regimens.

To establish the minimum level of change required to reliably distinguish relevant changes in PRO scores from random variations, we determined the retest reliability of each PRO (pain, fatigue, and function) in the discovery cohort (Table 2) based on data from months 12, 18, and 24. Function had the highest retest reliability (0.942), but all of the disease activity measures, including the PROs, showed relatively high values (>0.75).

Because responder analyses utilize individual changes rather than mean changes in a population, this method cannot accommodate fractional changes for integer-based scales (e.g., pain and fatigue). Accordingly, the values were rounded up to the nearest higher integer for responder analyses. On the basis of the calculated values, we conservatively determined that an individual improvement would require a decrease of ≥ 3 points for a therapeutic response in pain, a decrease of ≥ 4 points for a response in fatigue, and an increase of ≥ 16 points for a therapeutic response in function (FFbH). As in the previous study, we used a conservative DAS28- d_{crit} value of 1.8 to indicate a therapeutic response in disease activity, as our previous research showed that this value could be applied to all patients independent of their individual disease characteristics (4).

Table 1. Demographic data and disease characteristics of the discovery and treatment cohorts*

	Discovery cohort during stable treatment (n = 700)†				Treatment cohort (n = 2,788)	
	Prior to adalimumab initiation	Month 12‡	Month 18	Month 24	Prior to adalimumab initiation	Month 12
Age, years	54.3 ± 12.6	55.3 ± 12.6	55.8 ± 12.6	56.3 ± 12.6	54.6 ± 12.8	55.6 ± 12.8
Females, %	77.9	77.9	77.9	77.9	77.4	77.4
Disease duration, years	11.4 ± 9.5	12.4 ± 9.5	12.9 ± 9.5	13.4 ± 9.5	11.7 ± 9.2	12.7 ± 9.2
Tender joint count	9.7 ± 7.2	3.2 ± 4.8	3.5 ± 5.0	3.1 ± 4.8	12.6 ± 7.2	5.0 ± 6.0
Swollen joint count	7.2 ± 5.9	2.1 ± 3.5	2.1 ± 3.4	2.1 ± 3.6	9.7 ± 6.3	3.5 ± 4.5
CRP, mg/liter	17.3 ± 33.7	5.7 ± 11.0	6.2 ± 11.4	5.5 ± 10.3	32.2 ± 57.1	15.4 ± 33.6
ESR, mm/hour	29.8 ± 21.6	19.6 ± 17.1	19.6 ± 17.1	20.3 ± 18.2	33.8 ± 22.6	22.6 ± 18.6
PGA	5.9 ± 2.0	3.7 ± 2.0	3.7 ± 2.0	3.7 ± 2.1	6.6 ± 1.9	4.3 ± 2.1
MDGA	6.3 ± 1.9	2.7 ± 1.7	2.7 ± 1.8	2.6 ± 1.9	6.9 ± 1.6	3.4 ± 2.1
DAS28	5.2 ± 1.3	3.3 ± 1.3	3.3 ± 1.4	3.3 ± 1.4	5.9 ± 1.2	3.9 ± 1.5
% remaining function (FFbH score)	64.3 ± 22.2	74.4 ± 21.3	73.6 ± 22.3	73.4 ± 22.3	58.8 ± 23.1	68.3 ± 23.5
Pain	6.0 ± 2.2	3.5 ± 2.2	3.6 ± 2.2	3.5 ± 2.3	6.7 ± 2.0	4.1 ± 2.3
Fatigue	5.4 ± 2.6	3.5 ± 2.4	3.6 ± 2.5	3.5 ± 2.5	5.9 ± 2.6	4.0 ± 2.6

* Values are the mean ± SD unless indicated otherwise. Complete data were not available for all patients. Missing data were not imputed. CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; PGA = patient global assessment; MDGA = physician global assessment; DAS28 = Disease Activity Score in 28 joints; d_{crit} = critical difference; FFbH = Funktionsfragebogen Hannover.
† Period for determination of d_{crit} values.
‡ Month 12 was the time point for entry into the discovery cohort on the basis of stable disease activity.

Evaluation of PRO responses in patients achieving a DAS28-d_{crit} response. Having established d_{crit} values that exceeded the threshold of random variation using data from a discovery cohort with stable disease activity and treatment, we then applied these values to a population of patients with active RA. The applicability of data derived from a discovery cohort to a population with active disease was validated for the DAS28-d_{crit} in our previous publication (4). Statistically significant PRO responses were evaluated in a treatment cohort of 2,788 German patients initiating treatment with adalimumab in a non-interventional study. Compared with the discovery cohort, the treatment cohort had similar mean demographic characteristics at baseline (prior to adalimumab treatment initiation), but increased signs of disease activity as indicated by mean DAS28, swollen and tender joint counts, and inflammatory markers (Table 1). At month 12,

1,483 patients (53.2%) in the treatment cohort achieved a significant DAS28 response (improvement of at least 1.8 points from baseline), and 1,305 patients (46.8%) did not achieve a DAS28-d_{crit} response.

To provide insights into the association between PROs and objective responses of disease activity, we evaluated statistically significant PRO responses at month 12 in the subset of patients who achieved a therapeutic response as determined by a change in DAS28 of ≥1.8 from baseline at month 12. Of the 1,483 patients with a DAS28-d_{crit} response, 1,016 (68.5%) also showed a significant improvement in pain, 571 (38.5%) showed a significant improvement in fatigue, and 623 (42.0%) showed a significant improvement in function (Figure 1A and Figure 2). Overall, 1,145 (77.2%) reported a significant improvement in at least 1 PRO (pain, fatigue, or function), 338 (22.8%) did not achieve a significant PRO response for any outcome, and 336 (22.7%) showed improvement in all 3 PROs. In patients with only 1 PRO response, pain accounted for the vast majority of responses (20.9% versus 2.4% for fatigue and 4.8% for function). Similarly, in patients with 2 PRO responses, pain in combination with fatigue (11.9%) and pain in combination with function (13.0%) were the most common patterns of response; very few patients (1.6%) reported improvement in fatigue plus function in the absence of improvement in pain.

An alternative way of looking at PRO reporting patterns in patients with a therapeutic response at month 12 is presented in Figure 2. For each significant PRO, this figure illustrates the proportions of patients with a response in that PRO alone and with responses in multiple PROs. Although it was not uncommon for patients to experience a significant improvement in pain alone (20.9% of patients who reported an improvement in pain did not have corresponding

Table 2. Long-term retest reliability of PROs and the MDGA in the discovery cohort over a period of 1 year*

Outcome	No.	1-sided d _{crit}	Reliability
Function†	644	15.91	0.942
PGA	525	2.63	0.794
Fatigue	621	3.38	0.793
Pain	626	2.99	0.774
MDGA	676	2.45	0.757

* For each outcome, analyses were performed using 3 data sets from different time points (months 12, 18, and 24) and were limited to patients with a complete data set (data at months 12, 18, and 24) for the outcome being evaluated. PROs = patient-reported outcomes; MDGA = physician global assessment; PGA = patient global assessment.
† As determined by the Funktionsfragebogen Hannover score.

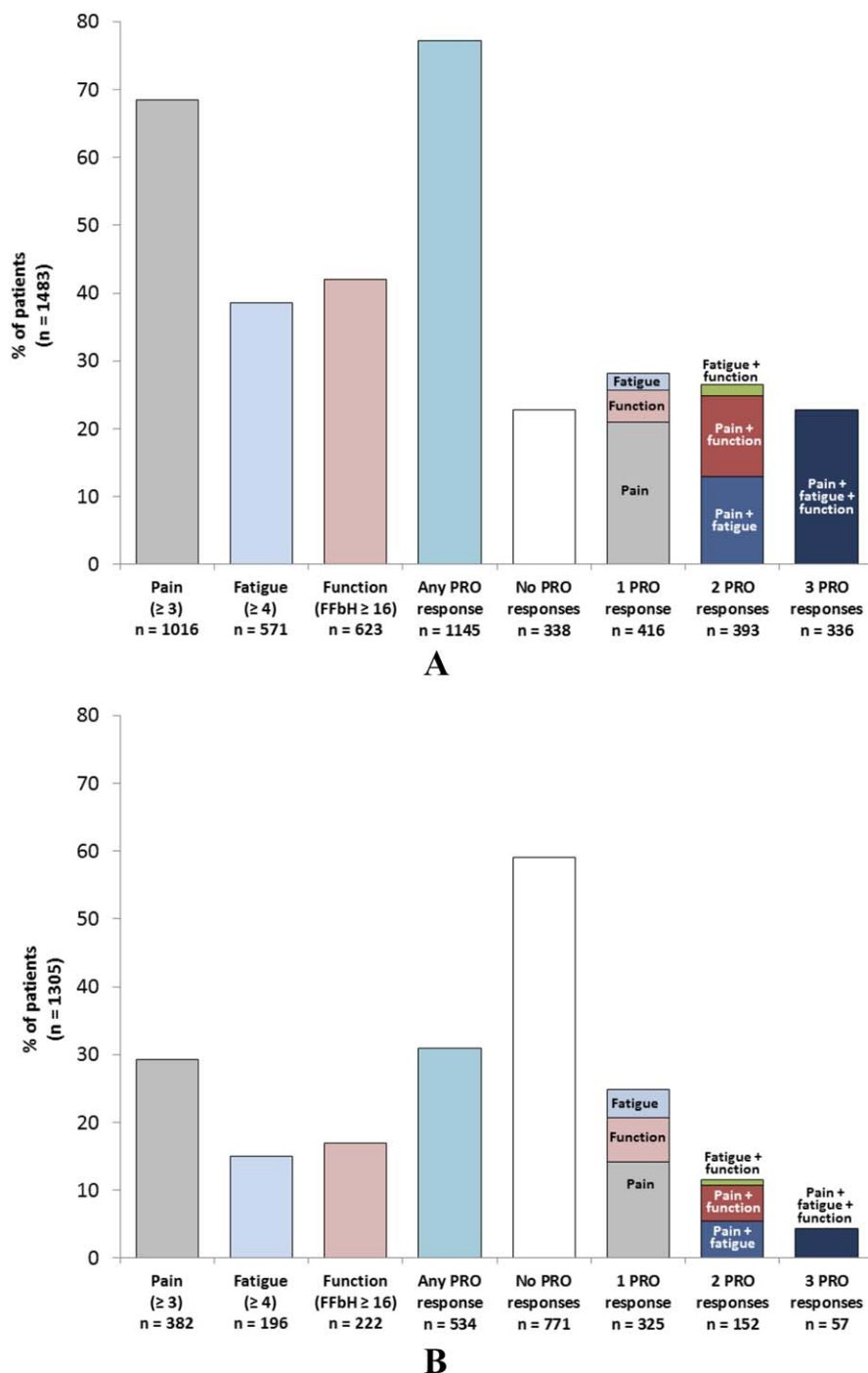


Figure 1. Significant responses in patient-reported outcomes (PROs) in the treatment cohort for **A**, patients with a significant individual response in Disease Activity Score in 28 joints (DAS28) (improvement ≥ 1.8) at month 12 and **B**, patients who did not achieve a significant DAS28 response at month 12. The designation of “1 PRO response” includes patients with a significant response in only 1 PRO, while “2 PRO responses” includes patients with a significant response in only 2 PROs. FFbH = Funktionsfragebogen Hannover questionnaire.

improvements in fatigue or function), it was quite rare for patients to report an improvement in fatigue alone or function alone (2.4% and 4.8%, respectively, of patients with improvements in those PROs). More commonly, patients who showed a significant therapeutic response in fatigue or function also had improvements in pain or in all 3 PROs.

Evaluation of PRO responses in patients not achieving a DAS28-d_{crit} response. We also evaluated PRO responses in patients who did not achieve a DAS28-d_{crit} response (n = 1,305) (Figure 1B). Of these patients, 382 (29.3%) showed a significant improvement in pain, 196 (15.0%) showed a significant improvement in fatigue, and 222

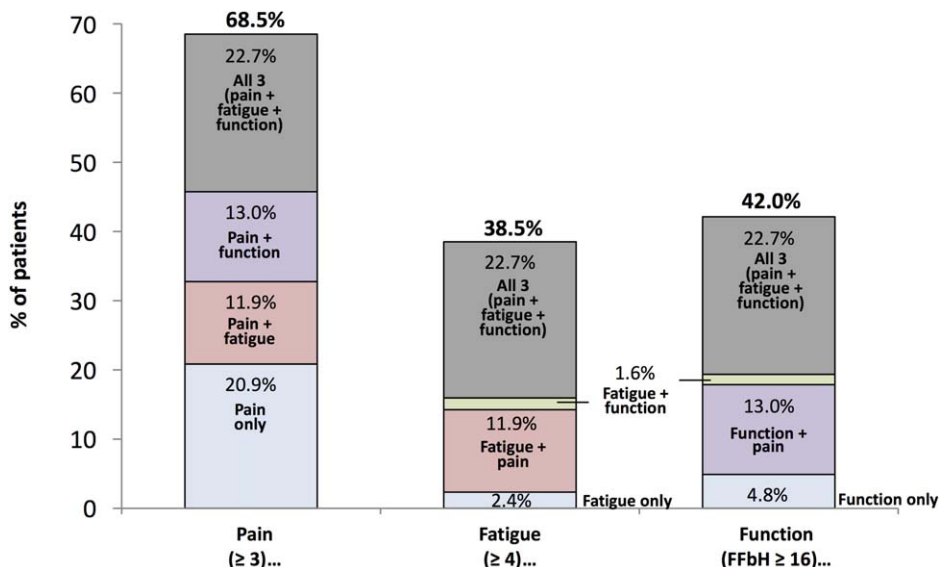


Figure 2. In-depth analysis of significant responses in patient-reported outcomes (PROs) in patients in the treatment cohort with a significant individual response in Disease Activity Score in 28 joints (improvement ≥ 1.8) at month 12 ($n = 1,483$). For each significant PRO response, the proportions of patients with a response in that PRO alone and with responses in multiple PROs are shown. FFbH = Funktionsfragebogen Hannover questionnaire.

(17.0%) showed a significant improvement in function (FFbH ≥ 16). Overall, 534 (40.9%) reported a significant improvement in at least 1 PRO (pain, fatigue, or function) and 57 (4.4%) showed improvement in all 3 PROs; 59.1% of patients did not experience any significant PRO response. Pain accounted for the majority of responses in patients with only 1 PRO response, and pain in combination with fatigue and pain in combination with function were the most common patterns of response for patients with 2 PRO responses.

Association between PROs and other disease activity assessments at baseline. We evaluated the association between PROs and other measures of disease activity in the treatment cohort at baseline (Table 3). As might be expected, the PRO that showed the highest correlation with DAS28 was PGA, which is a component of DAS28. DAS28 showed comparable correlations with function

and pain and a somewhat lower association with fatigue. Of the PROs evaluated, pain showed the highest correlation with PGA. The MDGA was moderately correlated with DAS28, PGA, and pain, but had lower associations with fatigue and function.

DISCUSSION

The findings reported here provide further validation for the use of statistically based individual improvements to evaluate changes in relevant outcomes during therapy and offer insights into PRO improvements that occur in patients with a DAS28-based therapeutic response. Our study also supports the critical role of pain in influencing patient-reported evaluations of disease activity, and suggests that improved pain management should be explored in patients who report high disease activity despite favorable changes in objective measures.

	DAS28	Function†	Fatigue	Pain	PGA	MDGA
DAS28	1.00	-0.41	0.29	0.42	0.49	0.52
Function†		1.00	-0.48	-0.51	-0.53	-0.31
Fatigue			1.00	0.58	0.59	0.28
Pain				1.00	0.79	0.41
PGA					1.00	0.46
MDGA						1.00

* The negative correlations observed with function are due to a difference in the direction of the scales; higher function scores indicate improved function, whereas for the other measures lower scores indicate improved outcomes. PROs = patient-reported outcomes; DAS28 = Disease Activity Score in 28 joints; PGA = patient global assessment; MDGA = physician global assessment.
 † As determined by the Funktionsfragebogen Hannover score.

All of the outcomes examined in this study showed excellent long-term test–retest reliability in RA patients on constant therapy. The PROs had reliability coefficients only slightly lower than the reliability coefficient of DAS28 (0.757–0.793 versus 0.826); these reliability coefficients compare favorably with values reported for other clinical parameters such as heart rate (0.44) or blood pressure (0.45–0.57) (15), indicating that PROs are valid and reproducible measures of patient status.

Despite the high reliability coefficients of PROs, a sizeable change was required to achieve a significant response outside the range of normal variation. For pain and fatigue, both of which were measured on scales ranging from 0 to 10, a decrease of 3 and 4 points, respectively, was required for a statistically significant improvement. For function, which was measured by the FFbH questionnaire on a scale ranging from 0 to 100, a change of 16 points was required; this was slightly lower than the d_{crit} value of 18 points determined previously in a smaller cohort of patients (4). Accordingly, these measures required a change from baseline representing approximately 20% to 40% of the full measurement scale. The level of change required for a significant improvement should be kept in mind by clinicians when evaluating alterations in PROs.

Other studies with different methodologies have reported various magnitudes for changes required for relevant changes in PROs in patients with RA. For pain measured on a 10-point visual analog scale (VAS), MCIDs of 0.5 (16) to 1.19 (17) and a short-term SDD of 1.63 (18) have been reported. For fatigue measured on a 10-point VAS, an MCID of 1.0 (19) and an SDD of 3.23 (18) have been reported. Studies using statistically based measures to determine values that exceed random variation, such as SDD or reliable change index, consistently report higher values than those that report MCID. This is best illustrated by a study of fatigue in patients with RA that reported both MCID and reliable change index (7). The MCID for improvement in fatigue in this study was 0.82 to 1.12, while the reliable change index was 3.47. The fact that the MCID falls well within the range of random variation as determined in this study calls into question the utility of MCID in assessing patient responses, and suggests that this methodology may not be optimal for long-term assessments of changes in PROs. Our methodology differs from those used to determine SDD or the reliability change index in that data for test–retest reliability are obtained over 1 year, rather than over hours or days as for these other statistically based tests. In clinical practice, patients with RA are seen every 3 to 6 months, so our methodology more closely reflects routine care.

In correlation analyses, PROs were found to be moderately correlated with DAS28 (Pearson's $r = 0.32$ – 0.42). The lack of a stronger correlation suggests that PROs reflect additional facets of disease activity that are not included in clinical evaluations. Other studies have also found that PROs are to some degree independent of clinical measures of disease activity (20,21). For instance, Kievit et al (20) reported that during routine treatment in patients with RA, self-reported PGA improved over time regardless of changes in DAS28, indicating that patients'

perceptions of health are not always in line with objective disease activity assessments.

In this study, pain was the PRO most highly correlated with PGA (Pearson's $r = 0.79$). Khan et al (22) also found that pain is the most important determinant of PGA; in their study, pain explained approximately 20% of the variance in PGA scores, while fatigue, the second most important determinant, explained approximately 6% of the variance. Pain is also the most significant factor contributing to discordancy between PGAs and MDGAs (22,23). Together, these findings suggest that pain should be given greater consideration when evaluating patients with RA, as this PRO has a significant impact on global well-being.

The key role played by pain was further revealed in our analyses of significant individual PRO responses occurring in patients with a therapeutic response at month 12 as assessed by DAS28- d_{crit} (improvement from baseline of at least 1.8 points) (4). More than two-thirds of patients with a DAS28 response also reported a significant improvement in pain, either alone or in combination with another PRO. In contrast, fewer than 40% of patients experienced improvements in fatigue or function; improvements in those PROs were typically accompanied by an improvement in pain as well.

In the present study, fatigue was not as dominant a factor as pain, although it was identified as the most critical PRO in a study by Minnock et al (24). Our study suggests that fatigue is not well represented in physician-based evaluations, such as the DAS28 or MDGA, but is an important component of a patient's self-perceived well-being as indicated by its correlation with PGA scores. A recent study found that the fatigue experienced by patients with RA is associated with RA disease activity, poor sleep quality, depression, and obesity (25); accordingly, improvements in disease activity alone may not be sufficient to decrease a patient's overall fatigue. Of the 3 PROs we evaluated, fatigue was the least likely to improve. Other studies have also found that fatigue is resistant to improvement during therapy (26,27). Nevertheless, 38.5% of patients who remained on therapy in our study did experience a significant response in fatigue after 12 months of adalimumab treatment, in agreement with other studies of the effect of anti-TNF agents, including adalimumab, on fatigue (28,29).

An important aspect of the current study is its relevance to the appropriate management of patients. In the cohort of patients initiating adalimumab therapy (treatment cohort), analyses of DAS28 and PRO responses allowed the identification of 3 groups of patients: 1) patients who were performing well with respect to both objective (DAS28) and subjective (PRO) measures, 2) patients who did not achieve a DAS28 response, but who had PRO responses, and 3) patients who achieved a DAS28 response, but not a PRO response. These 3 categories have significant implications with respect to treatment. For the first category, no change in RA treatment is required. In contrast, a change in RA treatment should be considered for patients in the second category. The third category poses the most difficult clinical dilemma. For these patients, additional analyses may be required to

differentiate between pain or fatigue related to RA and those due to other comorbidities, such as osteoarthritis or depression.

The findings reported here provide further support for the use of statistically derived, individual improvement criteria in the assessment of therapeutic response. As reported previously (4), this measure of response is clearly distinguishable from random variations in disease activity and is robust over time. Approximately three-fourths of patients with a DAS28 response had a significant improvement in pain, function, or fatigue, and more than 20% had a response in all 3 PROs, indicating that therapeutic responses as assessed by the DAS28- d_{crit} are associated with improvements in PROs in the majority of patients. In contrast, of the patients who did not achieve a DAS28- d_{crit} response at month 12, 59% had no significant PRO responses, and significant improvements in all 3 PROs occurred in only 4%.

Our findings may aid clinicians in informing patients of the typical outcomes achieved during anti-TNF therapy and in managing treatment expectations. In addition, the results of this study suggest that clinicians should consider evaluating PROs in patients who report disease activity despite objective improvements, or introducing PRO-based assessments, such as the Routine Assessment of Patient Index Data 3, which includes patient-reported physical function, pain, and global assessment (30) into routine clinical practice. Particular attention should be paid to self-reported pain, as pain appears to play a key role in patient perceptions of overall well-being.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Tony had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Scharbatke, Behrens, Schmalzing, Koehm, Greger, Gnann, Burkhardt, Tony.

Acquisition of data. Scharbatke, Behrens, Schmalzing, Koehm, Burkhardt, Tony.

Analysis and interpretation of data. Scharbatke, Behrens, Schmalzing, Koehm, Greger, Gnann, Burkhardt, Tony.

ROLE OF THE STUDY SPONSOR

AbbVie Deutschland GmbH & Co. KG (formerly Abbott GmbH & Co. KG) provided funding for this study and was involved in the study design, data collection, data analysis, review of the manuscript for scientific integrity, and approval of the content of the submitted manuscript. The study sponsor did not influence data interpretation. Sharon L. Cross, PhD, provided medical writing services on behalf of CIRI (Frankfurt am Main, Germany) under contract with AbbVie Deutschland GmbH & Co. KG.

ADDITIONAL DISCLOSURE

Author Gnann is a paid consultant for AbbVie Deutschland GmbH & Co. KG.

REFERENCES

- Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. *Rheumatology (Oxford)* 2004;43:704–11.
- Pincus T, Castrejón, Yazici Y. Documenting the value of care for rheumatoid arthritis, analogous to hypertension, diabetes, and hyperlipidemia: is control of individual patient self-report measures of global estimate and physical function more valuable than laboratory tests, radiographs, indices, or remission criteria? *J Rheumatol* 2013;40:1469–74.
- Kleinert S, Tony HP, Krause A, Feuchtenberger M, Wassenberg S, Richter C, et al. Impact of patient and disease characteristics on therapeutic success during adalimumab treatment of patients with rheumatoid arthritis: data from a German noninterventional observational study. *Rheumatol Int* 2012;32:2759–67.
- Behrens F, Tony HP, Alten R, Kleinert S, Scharbatke EC, Köhm M, et al. Development and validation of a new Disease Activity Score in 28 joints–based treatment response criterion for rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2013;65:1608–16.
- Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther* 2012;20:160–6.
- Auleley GR, Benbouazza K, Spoorenberg A, Collantes E, Hajjaj-Hassouni N, van der Heijde D, et al. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. *Arthritis Rheum* 2002;15:47:582–7.
- Khanna D, Pope J, Khanna PP, Malone M, Samedí N, Norrie D, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008; 35:2339–43.
- Prevo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44–8.
- Vander Cruyssen B, van Looy S, Wyns B, Westhovens R, Durez P, van den Bosch F, et al. DAS28 best reflects the physician's clinical judgment of response to infliximab therapy in rheumatoid arthritis patients: validation of the DAS28 score in patients under infliximab treatment. *Arthritis Res Ther* 2005;7:R1063–71.
- Raspe HH, Hagedorn U, Kohlmann T, Mattussek S. Der Funktionsfragebogen Hannover (FFbH): Ein Instrument zur Funktionsdiagnostik bei polyartikulären Gelenkerkrankungen. In: Siegrist J, ed. *Wohnortnahe Betreuung Rheumakrankter. Ergebnisse sozialwissenschaftlicher Evaluation eines Modellversuches*. Stuttgart (Germany): Schattauer; 1990. pp. 164–82.
- Lautenschläger J, Mau W, Kohlmann T, Raspe HH, Struve F, Brückle W, et al. Comparative evaluation of a German version of the Health Assessment Questionnaire and the Hannover Functional Capacity Questionnaire. *Z Rheumatol* 1997;56: 144–55.
- Westhoff G, Listing J, Zink A. Loss of physical independence in rheumatoid arthritis; interview data from a representative sample of patients in rheumatologic care. *Arthritis Care Res* 2000;13:11–22.
- Lienert GA, Raatz U. *Testaufbau und testanalyse (test construction and test analysis)* Weinheim (Germany): Beltz/ Psychologie Verlagsunion; 1998.
- Winer BJ, Brown DR, Michels KM. *Statistical principles in experimental design*. 3rd ed. New York: McGraw-Hill; 1991.
- Gerin W, Christenfeld N, Pieper C, DeRafael DA, Su O, Stroessner SJ, et al. The generalizability of cardiovascular responses across settings. *J Psychosom Res* 1998;44:209–18.
- Wolfe F, Michaud K. Assessment of pain in rheumatoid arthritis: minimal clinically significant difference, predictors, and the effect of anti-tumor necrosis factor therapy. *J Rheumatol* 2007;34:1674–83.

17. Pope JE, Khanna D, Norrie D, Ouimet JM. Minimally important difference for the health assessment questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *J Rheumatol* 2009;36:254–9.
18. Kvien TK, Mowinckel P, Heiberg T, Dammann KL, Dale Ø, Annerud GJ, et al. Performance of health status measures with a pen based personal digital assistant. *Ann Rheum Dis* 2005;64:1480–4.
19. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280–9.
20. Kievit W, Welsing PM, Adang EM, Eijsbouts AM, Krabbe PF, van Riel PL. Comment on the use of self-reporting instrument to assess patients with rheumatoid arthritis: the longitudinal association between the DAS28 and the VAS general health. *Arthritis Rheum* 2006;55:745–50.
21. Ward MM, Guthrie LC, Alba MI. Brief report: rheumatoid arthritis response criteria and patient-reported improvement in arthritis activity: is an American College of Rheumatology twenty percent response meaningful to patients? *Arthritis Rheumatol* 2014;66:2339–43.
22. Khan NA, Spencer HJ, Abda E, Aggarwal A, Alten R, Ancuta C, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. *Arthritis Care Res (Hoboken)* 2012;64:206–14.
23. Castrejón I, Yazici Y, Samuels J, Luta G, Pincus T. Discordance of global estimates by patients and their physicians in usual care of many rheumatic diseases: association with 5 scores on a Multidimensional Health Assessment Questionnaire (MDHAQ) that are not found on the Health Assessment Questionnaire (HAQ). *Arthritis Care Res (Hoboken)* 2014;66:934–42.
24. Minnock P, Kirwan J, Bresnihan B. Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. *Rheumatology (Oxford)* 2009;48:1533–6.
25. Katz P, Margaretten M, Trupin L, Schmajuk G, Yazdany J, Yelin E. Role of sleep disturbance, depression, obesity, and physical inactivity in fatigue in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2016;68:81–90.
26. Curtis JR, Shan Y, Harrold L, Zhang J, Greenberg JD, Reed GW. Patient perspectives on achieving treat-to-target goals: a critical examination of patient-reported outcomes. *Arthritis Care Res (Hoboken)* 2013;65:10:1707–12.
27. Campbell RC, Batley M, Hammond A, Ibrahim F, Kingsley G, Scott DL. The impact of disease activity, pain, disability and treatments on fatigue in established rheumatoid arthritis. *Clin Rheumatol* 2012;31:717–22.
28. Yount S, Sorensen MV, Cella D, Sengupta N, Grober J, Chartash EK. Adalimumab plus methotrexate or standard therapy is more effective than methotrexate or standard therapies alone in the treatment of fatigue in patients with active, inadequately treated rheumatoid arthritis. *Clin Exp Rheumatol* 2007;25:838–46.
29. Minnock P, McGee G, Bresnihan B, FitzGerald O, Veale DJ. How much is fatigue explained by standard clinical characteristics of disease activity in patients with inflammatory arthritis? A longitudinal study. *Arthritis Care Res (Hoboken)* 2014;66:1597–603.
30. Pincus T, Furer V, Keystone E, Yazici Y, Bergman MJ, Luijckens K. RAPID3 (Routine Assessment of Patient Index Data 3) severity categories and response criteria: similar results to DAS28 (Disease Activity Score) and CDAI (Clinical Disease Activity Index) in the RAPID 1 (Rheumatoid Arthritis Prevention of Structural Damage) clinical trial of certolizumab pegol. *Arthritis Care Res (Hoboken)* 2011;63:1142–9.