

RMD Open

Rheumatic &
Musculoskeletal
Diseases

EXTENDED REPORT

Psychometric properties of three single-item pain scales in patients with rheumatoid arthritis seen during routine clinical care: a comparative perspective on construct validity, reproducibility and internal responsiveness

Melanie Sendlbeck, Elizabeth G Araujo, Georg Schett, Matthias Englbrecht

To cite: Sendlbeck M, Araujo EG, Schett G, *et al*. Psychometric properties of three single-item pain scales in patients with rheumatoid arthritis seen during routine clinical care: a comparative perspective on construct validity, reproducibility and internal responsiveness. *RMD Open* 2015;1:e000140. doi:10.1136/rmdopen-2015-000140

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/rmdopen-2015-000140>).

Received 23 June 2015
Accepted 20 October 2015



CrossMark

Department of Internal Medicine 3 and Institute of Clinical Immunology, University of Erlangen-Nuremberg, Erlangen, Germany

Correspondence to
Dr Matthias Englbrecht;
matthias.englbrecht@uk-erlangen.de

ABSTRACT

Objective: To investigate the construct validity, reproducibility (ie, retest reliability) and internal responsiveness to treatment change of common single-item scales measuring overall pain in patients with rheumatoid arthritis (RA) and to investigate the corresponding effect of common pain-related comorbidities and medical consultation on these outcomes.

Methods: 236 patients with RA completed a set of questionnaires including a visual analogue scale (VAS), a numerical rating scale (NRS) and a verbal rating scale (VRS) measuring overall pain before and immediately after routine medical consultation as well as 1 week after the patient's visit. Construct validity and retest reliability were evaluated using the Bravais-Pearson correlation while standardised response means (SRM) were calculated for evaluating internal responsiveness. Differences in the perception of pain were calculated using dependent samples t-tests.

Results: In the total sample, construct validity was good across all three time points (convergent validity of pain scales: $r_{T1-T3}=0.82-0.92$, $p<0.001$; discriminant validity as correlation of pain scales with age: $r_{age}=0.01-0.16$, $p>0.05$). In patients maintaining antirheumatic treatment, retest reliability of pain scales was confirmed for all scales and across time points ($r_{VAS}=0.82-0.95$, $r_{NRS}=0.89-0.98$, $r_{VRS}=0.80-0.90$, $p<0.001$), while the internal responsiveness of scales to a change in treatment was low across all scales (SRM=0.08–0.21). The VAS especially suggested a change in pain perception after medical consultation in patients maintaining therapy.

Conclusions: The VAS, NRS and VRS are valid and retest reliable in an outpatient clinical practice setting. The low pain scales' internal responsiveness to treatment change is likely to be due to the short follow-up period. Patients with RA maintaining antirheumatic therapy seem to experience less pain after medical consultation.

Key messages

What is already known about this subject?

- The visual analogue scale, the numerical rating scale and the verbal rating scale measuring overall pain in arthritis are frequently used in routine clinical care as well as in clinical trials in order to document the patient's current extent of pain.

What does this study add?

- With this study, we evaluated a psychometric profile of these three single-item measures including construct validity, reproducibility and internal responsiveness while accounting for chronic pain and/or osteoarthritis.
- In patients on stable antirheumatic treatment but not in patients changing treatment, pain levels on a visual analogue scale decreased after medical consultation compared to the ratings before consultation.

How might this impact on clinical practice?

- With the scales showing valid and reliable overall results, the choice of one of these instruments for measuring pain depends on the purpose and the corresponding clinical situation including individual preferences.
- The aspect of decreasing pain during medical consultation shown on a visual analogue scale might hint at the critical role of physician-patient interaction and has to be investigated in more detail in further studies.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory joint disease characterised by synovitis, joint destruction and impaired physical function.¹ Beside symptoms such as

fatigue, impaired sleep quality and physical functioning, pain has recently been shown to have the largest impact of disease symptoms on life in patients with RA.² The Encyclopedia of Pain refers to it as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage”.³ If present, it substantially reduces quality of life, consequently affects daily routines, and poses a challenge to emotional, physical as well as social resources.

From a rheumatological perspective, accurate measurement of pain in patients with RA is essential in clinical practice as well as in clinical trials requiring the use of valid and reliable methods of pain assessment. Currently, there is a variety of single-item and composite pain measures available focusing on different aspects of pain, such as frequency, severity, duration or anatomical location.^{4–5} The visual analogue scale (VAS), the numerical rating scale (NRS) and the verbal rating scale (VRS)—the latter is sometimes also referred to as the Likert scale—as single-item tools measuring overall pain are recommended and commonly used in clinical practice for pain assessment in inflammatory arthritis.⁴ Their brevity has also led to their application in non-clinical settings and approaches to aim for national standards.⁶ In rheumatology, however, a simple but important question remained unanswered: are these frequently applied single-item scales sufficiently valid and reliable in the context of a real-life clinical patient sample? Yet, the corresponding literature has not been able to answer this question satisfactorily: Either the scope of the articles focused on single psychometric aspects or the patient sample encompassed a variety of rheumatic (and potentially even non-inflammatory) diagnoses.^{7–16} Accordingly, a direct comparison of these findings is not feasible and a complete psychometric profile of single-item pain scales (including validity, reliability and responsiveness) in one sample is still missing. Corresponding results from a systematic literature review reflect this fact very well.¹⁷

To fill the remaining gap, the objective of this paper is to investigate the psychometric properties of these tools in a sample of patients with RA. Additionally, we were interested in investigating whether routine medical consultation might lead to changes in patient-reported pain before and after consultation as recent literature on physician-patient interaction may suggest.^{18–19} Furthermore, we also accounted for limitations of physical functioning and common pain-related comorbidities (ie, osteoarthritis and/or chronic pain) to see whether they interfere with the scales' psychometric properties.

PATIENTS AND METHODS

Patient sample and study design

After the approval of the local institutional review board, 300 patients with RA from the outpatient clinic of the

Department of Internal Medicine 3 at the University of Erlangen-Nuremberg were asked to voluntarily participate in a survey focusing on individual experience with RA. Patients were consecutively asked for participation with no limitations due to disease activity, medical history or other characteristics. They were enrolled from December 2011 to March 2013. Two hundred and thirty-six patients agreed to participate and signed an informed consent form. Patients completed a set of questionnaires consisting of (1) the single-item VAS, NRS and VRS asking for the extent of overall pain in the past 7 days, (2) the Health Assessment Questionnaire—Disability Index (HAQ-DI)²⁰ and (3) three items of the Rheumatoid Arthritis Disease Activity Index (RADAI).²¹ In order to distract the patients' attention from the last rating on one of the scales of interest, the pain scales were spread across the questionnaire. All participants were asked to complete this set of questionnaires three times: Before and immediately after medical consultation (T1 & T2) as well as 1 week after this visit at home (ie, days 6–13) (T3). Each of the three questionnaires comprised 48 items and had identical item order. The short follow-up period was favoured over longer periods in order to be able to reflect changes in pain ratings after dose escalation, tapering or start with glucocorticoid or non-steroidal anti-inflammatory drug therapy. Additional information taken from the patient's record included: demographic background (ie, age and sex), disease-related variables such as duration of disease, presence of autoantibodies, the Disease Activity Score 28 (DAS28),^{22–23} changes in antirheumatic therapy potentially influencing pain and frequent pain-related comorbidities (ie, chronic pain and/or osteoarthritis). The latter information was obtained in order to be able to investigate whether psychometric results are different in patients with RA with and without osteoarthritis and/or chronic pain. A treatment change was defined as any change in dosage or substance with respect to non-steroidal anti-inflammatory drugs, biological or conventional disease-modifying antirheumatic drugs (DMARDs) or glucocorticoids.

The wording and the end points of the three pain scales were identical (figure 1). However, owing to the different nature of the scales, the values assigned to individual pain ratings were different: both the VAS and the NRS ranged from 0 (=no pain) to 10 (=worst pain imaginable); while the VAS included decimals, the NRS merely referred to integers. The VRS ranged from 0 (=no pain) to 4 (=worst pain imaginable), which in the context of the data analysis was treated as having interval properties according to Lord and Novick's landmark contribution on statistical theory of mental test scores.²⁴ In order to avoid any misunderstanding in view of the VRS: just as the VAS and the NRS, the VRS was a (Likert scaled) written item which was not asked orally as its usual denomination frequently used in literature might suggest. With some patients either not indicating a date of completion or completing the third questionnaire

as used in the present study (English translation)

Visual Analogue Scale (VAS) – arthritis related				
How much pain did you have in the last 7 days?				
Please pick a point on the given line!				
no pain	————— worst pain imaginable			
Numerical Rating Scale (NRS) – arthritis related				
How much pain did you have in the last 7 days?				
Please mark the appropriate number!				
no pain	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/> 10			
	worst pain imaginable			
Verbal Rating Scale (VRS) – arthritis related				
How much pain did you have in the last 7 days?				
Please mark the appropriate statement!				
no pain	slight pain	moderate pain	severe pain	worst pain imaginable
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Annotation: The VAS is a 10cm unnumbered horizontal line ranging from 0cm (= no pain) to 10cm (= worst pain imaginable). This way, patients have the possibility to score with high precision. Usually, in the paper-based version of the scale, a ruler is used in order to measure the distance from 0 to the patient's mark on the line. The NRS in contrast is segmented in numbers from 0 (= no pain) to 10 (= worst imaginable pain) which implies 11 scoring possibilities. On the VRS, patients were asked to choose one out of five categories for pain evaluation: no pain, slight pain, moderate pain, severe pain or worst imaginable pain. The German wording for the corresponding pain items with respect to arthritic pain was: "Wie würden Sie die Stärke Ihrer Schmerzen insgesamt in den letzten 7 Tagen einschätzen?"</p>				

Figure 1 The visual analogue scale, the numerical rating scale and the verbal rating scale as used in the present study (English translation).

after the regular follow-up period was over, we computed two alternative data analyses for the last time point: One included the complete sample of patients while the other was limited to participants who had completed the last questionnaire during days 6–13 after consultation (N=111). Accordingly, there are two sets of results available for the last time point (T3), whereas results for the first two time points always refer to the total sample. Missing values were not imputed by any means.

Evaluation of psychometric properties and corresponding statistical data analysis

In order to evaluate the psychometric properties of the VAS, the NRS and the VRS, we evaluated convergent as well as discriminant validity (ie, two subsets of construct validity) by means of a multitrait-multimethod approach referring to the ideas of Campbell and Fiske.²⁵ Following this idea, the Bravais-Pearson correlation of indicators measuring the same construct (ie, convergent validity of scales measuring overall pain) were supposed to be $r \geq 0.45$ ²⁶ and substantially larger than the

correlation of indicators measuring different constructs such as correlations of pain scales with age or disease duration which was supposed to reflect discriminant validity. In investigations involving pain measures, the potential interference with physical functioning frequently is a matter of concern. Therefore, we also calculated partial correlation coefficients for convergent validity as well as retest-reliability accounting for the influence of the HAQ-DI at corresponding time points.

Patients participating in this study were included in two subgroups depending on whether or not they required a change in antirheumatic therapy. In patients maintaining antirheumatic treatment, we investigated the retest reliability as an indicator of reproducibility between all pairs of time points for each of the three pain scales by Bravais-Pearson correlation coefficients. Coefficients of $r \geq 0.70$ were considered reliable.²⁶ In patients undergoing a change in antirheumatic therapy, for example, owing to side effects or lack of efficacy, we investigated the internal responsiveness of the pain scales to changes of antirheumatic treatment by

calculating standardised response means (SRMs) between T1 and T3 as well as between T2 and T3. Corresponding SRMs were computed by dividing the mean change by the SD of the change scores. We favoured this measure of responsiveness over an additional health transition item at the last follow-up (measuring external responsiveness) as such an item would have compromised our intention to keep all sets of questionnaires identical. In the context of studies validating patient-reported outcomes, an SRM exceeding 0.8 is considered large.²⁷ Differences in the perception of pain before and after routine medical consultation were investigated via a t-test for dependent samples including a report of the corresponding effect size (r).

A subsequent analysis investigated these characteristics in patients with RA with and without comorbidity of osteoarthritis and/or chronic pain. In this analysis, we used the same coefficients that were applied for the total sample, whereas Spearman correlation coefficients (r_s) or Wilcoxon signed-rank tests were calculated in comparisons including smaller subgroups ($N < 30$). In this subsample, partial correlation coefficients and measures of internal responsiveness were omitted due to sample size limitations. Statistical calculations were computed using IBM SPSS V.21, while all inferential tests were two tailed. Descriptive results in the text are presented as mean \pm SD if not stated otherwise. Effect sizes are presented as r or r^2 , respectively.

RESULTS

Patient sample characteristics

Data from 236 patients with RA fulfilling the 2010 American College of Rheumatology/European League Against Rheumatism classification criteria²⁸ were included in the analysis. Two hundred and ten patients with RA completed the set of questionnaires at all three time points. Patients had an average age and disease duration of 56.87 \pm 13.39 and 9.11 \pm 8.88 years, respectively. In total 73.7% ($N=174$) were female and 25.6% ($N=60$) of all patients underwent a change in antirheumatic treatment on the day of medical consultation. In total 15.3% ($N=36$) of the participants had a known history of osteoarthritis ($N=31$) and/or chronic pain ($N=6$) according to their clinical record. From those patients with available DAS28 ($N=208$), 38.9% ($N=81$) were in clinical remission ($DAS28 < 2.6$), 46.6% ($N=97$) had low-to-moderate disease activity ($2.6 \leq DAS28 < 5.1$), and 14.4% ($N=30$) had a $DAS28 > 5.1$ indicating high disease activity. On average, participants reported some difficulties in daily life due to physical limitations with respect to HAQ-DI scores (0.89 \pm 0.75). Pearson correlations of the pain scales with DAS28 ranged from $r=0.63$ to 0.71 (all $p < 0.001$), indicating the relevance of pain. Mean pain ratings before consultation were 3.44 \pm 2.61 cm on the VAS, 3.35 \pm 2.49 on the NRS and 1.50 \pm 0.91 on the VRS (table 1).

Construct validity in terms of convergent and discriminant validity

Bravais-Pearson correlation coefficients demonstrated good convergent validity for all pain scales at all time points with an amount of shared variance of at least 67.2% ($r_{T1}=0.82-0.89$; $r_{T2}=0.86-0.91$; $r_{T3}=0.86-0.92$, $r^2 \geq 0.67$, all $p < 0.001$). Even after control of the influence of physical functioning (HAQ-DI) on pain ratings by partial correlational analysis, relations among pain scales remained substantial with at least 51.8% of shared variance ($r_{T1}=0.72-0.83$; $r_{T2}=0.77-0.84$; $r_{T3}=0.78-0.86$, $r^2 \geq 0.52$, all $p < 0.001$).

With respect to the discriminant validity of the pain scales, absolute values of correlations with age ($|r_{age}|=0.01-0.12$, $r^2 \leq 0.01$, $p \geq 0.096$) and disease duration ($|r_{duration}|=0.09-0.16$, $r^2 \leq 0.03$, $p \geq 0.021$) showed substantially smaller coefficients with a maximum proportion of shared variance of only 2.6%.

In patients completing the last questionnaire during days 6–13 after consultation, the aforementioned pattern of results remained comparable (convergent validity: $r_{T3}=0.85-0.91$, $r^2 \geq 0.72$, $p < 0.001$; after accounting for the influence of physical functioning: $r_{T3}=0.75-0.85$, $r^2 \geq 0.56$, $p < 0.001$; discriminant validity: $|r_{age}|=0.01-0.16$, $r^2 \leq 0.03$, $p \geq 0.088$; $|r_{duration}|=0.13-0.22$, $r^2 \leq 0.05$, $p \geq 0.027$). With convergent coefficients being substantially larger than their discriminant counterparts for all of the three pain scales, construct validity may be assumed⁶ (figure 2 for further information).

Retest-reliability and responsiveness

In patients maintaining antirheumatic therapy, Bravais-Pearson coefficients for retest-reliability indicated sufficient reliability in the total sample ($r_{VAS}=0.82-0.95$, $r_{NRS}=0.89-0.98$, $r_{VRS}=0.80-0.90$, $r^2 \geq 0.64$, all $p < 0.001$) as well as in patients completing the last questionnaire during days 6–13 after the visit ($r_{VAS}=0.77-0.79$, $r_{NRS}=0.89-0.90$, $r_{VRS}=0.78-0.83$, $r^2 \geq 0.59$, all $p < 0.001$). Accordingly, the amount of shared variance on one pain scale between any pair of time points was 59.3% at least (for detailed information on coefficients, table 2). A control of the influence of physical functioning yielded a similar pattern of results for the total sample ($r_{VAS}=0.74-0.95$, $r_{NRS}=0.82-0.96$, $r_{VRS}=0.70-0.85$, $r^2 \geq 0.49$, all $p < 0.001$) as well as for the days 6–13 subgroup ($r_{VAS}=0.66-0.70$, $r_{NRS}=0.82-0.84$, $r_{VRS}=0.68-0.75$, $r^2 \geq 0.44$, all $p < 0.001$), although some coefficients were found to be slightly below 0.7 in the subsample.

SRMs between T1 and T3 as well as between T2 and T3 for patients undergoing antirheumatic treatment change were low ($SRM_{VAS}=0.14-0.21$, $SRM_{NRS}=0.08-0.10$, and $SRM_{VRS}=0.14-0.17$). In patients completing the last questionnaire between days 6 and 13 and undergoing treatment change, internal responsiveness was not observable ($SRM_{VAS}=-0.09-0.06$, $SRM_{NRS}=0.00-0.04$ and $SRM_{VRS}=0.00$). We obtained similar results for internal responsiveness of the three pain scales when limiting the analysis to patients treated with drugs

Table 1 Descriptive sample characteristics

	N	Minimum	Maximum	Mean	SD	95% CI of mean
Age	236	21	84	56.87	13.39	(55.16 to 58.59)
Duration of disease (years)	236	0	44	9.11	8.88	(7.97 to 10.24)
Sex (N female/N male)	236	174 (73.7%)/62 (26.3%)				
Rheumatoid factor (N positive/N negative)	234	180 (76.9%)/54 (23.1%)				
Cyclic citrullinated peptides antibodies (N positive/N negative)	236	152 (64.4%)/84 (35.6%)				
Methotrexate therapy	110					
Glucocorticoid therapy	93					
Biological therapy	131					
Change of medication at visit (N yes/N no)	234	60 (25.6%)/174 (74.4%)				
HAQ-DI (before consultation)	231	0	2.88	0.89	0.75	(0.79 to 0.98)
HAQ-DI (after consultation)	230	0	3.00	0.88	0.76	(0.78 to 0.98)
HAQ-DI (follow-up, total sample)	208	0	3.00	0.85	0.76	(0.74 to 0.95)
HAQ-DI (follow-up, days 6–13)	111	0	2.63	0.69	0.69	(0.56 to 0.82)
DAS28 (composite index)	208	0	7.61	3.33	1.54	(3.12 to 3.54)
VAS pain (before consultation)	213	0	10.0	3.44	2.61	(3.08 to 3.79)
VAS pain (after consultation)	209	0	9.9	3.27	2.63	(2.91 to 3.63)
VAS pain (follow-up, total sample)	192	0	9.1	2.98	2.37	(2.64 to 3.32)
VAS pain (follow-up, days 6–13)	104	0	9.0	2.55	2.17	(2.13 to 2.97)
NRS pain (before consultation)	226	0	10	3.35	2.49	(3.03 to 3.68)
NRS pain (after consultation)	228	0	9	3.32	2.48	(2.99 to 3.64)
NRS pain (follow-up, total sample)	206	0	9	3.24	2.38	(2.91 to 3.57)
NRS pain (follow-up, days 6–13)	110	0	9	2.84	2.27	(2.41 to 3.27)
VRS pain (before consultation)	223	0	3	1.50	0.91	(1.38 to 1.62)
VRS pain (after consultation)	224	0	3	1.45	0.91	(1.33 to 1.57)
VRS pain (follow-up, total sample)	203	0	4	1.38	0.90	(1.25 to 1.50)
VRS pain (follow-up, days 6–13)	107	0	3	1.20	0.86	(1.03 to 1.36)

Annotation: [table 1](#) shows descriptive information about the characteristics of the patient sample, including number of patients with data available (N), minimum and maximum values, arithmetic mean (mean), SD and 95% CI of the arithmetic mean (95% CI of mean). For categorical variables, the size of the corresponding subgroups is presented. DAS28, Disease Activity Score 28; HAQ-DI, Health Assessment Questionnaire-Disability Index; NRS, numerical rating scale; VAS, visual analogue scale; VRS, verbal rating scale.

	VAS pain (T1)	NRS pain (T1)	VRS pain (T1)	VAS pain (T2)	NRS pain (T2)	VRS pain (T2)	VAS pain (T3)	NRS pain (T3)	VRS pain (T3)	VAS pain (T3: 6-13)	NRS pain (T3: 6-13)	VRS pain (T3: 6-13)
	r	r	r	r	r	r	r	r	r	r	r	r
	N	N	N	N	N	N	N	N	N	N	N	N
VAS pain (T1)		0.83*** 192	0.72*** 192									
NRS pain (T1)	0.89*** 205		0.80*** 192									
VRS pain (T1)	0.82*** 202	0.88*** 215										
VAS pain (T2)					0.84*** 189	0.77*** 189						
NRS pain (T2)				0.91*** 202		0.82*** 189						
VRS pain (T2)				0.86*** 198	0.90*** 220							
VAS pain (T3)								0.86*** 181	0.78*** 181			
NRS pain (T3)							0.92*** 190		0.82*** 181			
VRS pain (T3)							0.86*** 186	0.89*** 200				
VAS pain (T3: 6-13)											0.85*** 97	0.75*** 97
NRS pain (T3: 6-13)										0.91*** 103		0.82*** 97
VRS pain (T3: 6-13)										0.85*** 100	0.89*** 107	
Age	0.06 213	0.01 226	0.00 223	0.09 209	-0.01 224	0.02 228	0.09 192	0.12 206	0.07 203	0.07 104	0.16 110	0.10 107
Disease duration (years)	0.09 213	0.10 226	0.10 223	0.13 209	0.14 [†] 224	0.11 228	0.14 192	0.11 206	0.16 [†] 203	0.18 104	0.14 110	0.19 [†] 107

Convergent validity after control of physical functioning
 ⇒ partial correlation of pain scales after control of HAQ-DI at corresponding timepoint

Convergent validity
 ⇒ bivariate Bravais-Pearson correlation of pain scales

Figure 2 Correlation coefficients for convergent and discriminant validity.

Table 2 Bravais-Pearson correlation coefficients for retest reliability

	VAS pain (T1)	VAS pain (T2)	VAS pain (T3)	VAS pain (T3: 6-13)	NRS pain (T1)	NRS pain (T2)	NRS pain (T3)	NRS pain (T3: 6-13)	VRS pain (T1)	VRS pain (T2)	VRS pain (T3)	VRS pain (T3: 6-13)
r	r	r	r	r	r	r	r	r	r	r	r	r
N	N	N	N	N	N	N	N	N	N	N	N	N
VAS pain (T1)	0.95 ^{***}	0.82 ^{***}	0.77 ^{***}									
	159	146	75									
VAS pain (T2)		0.86 ^{***}	0.79 ^{***}									
		144	75									
NRS pain (T1)				0.98 ^{***}			0.90 ^{***}	0.89 ^{***}				
				163			154	79				
NRS pain (T2)					0.90 ^{***}		0.89 ^{***}	0.90 ^{***}				
					156		156	80				
VRS pain (T1)									0.90 ^{***}		0.80 ^{***}	0.78 ^{***}
									158		149	76
VRS pain (T2)										0.80 ^{***}	0.80 ^{***}	0.83 ^{***}
										148	148	74

Correlation coefficients for each pain scale in view of retest reliability. Retest reliability may be assumed with coefficients of $r \geq 0.70$. Significant correlations are flagged, whereas $***p < 0.001$. NRS, numerical rating scale; VAS, visual analogue scale; VRS, verbal rating scale.

typically prescribed for pain relief (ie, glucocorticoids, non-steroidal anti-inflammatory drugs or opioids) and mostly receiving concomitant conventional or biological DMARDs (total sample: SRM=0.06–0.21, days 6–13 subgroup: SRM=-0.17–0.06). However, the corresponding subgroups were small ($N < 30$), which is why results should be considered cautiously.

Changes in patient-reported pain before and after medical consultation

Interestingly, patient-reported pain before medical consultation differed from the pain ratings on a VAS and NRS immediately after consultation in patients maintaining antirheumatic therapy, whereas the corresponding effect was small: $t_{VAS(T1-T2)}(158)=2.93$, $p=0.004$, $|r|=0.23$; $t_{NRS(T1-T2)}(162)=2.29$, $p=0.023$, $|r|=0.18$. These results suggest a slight improvement of pain during medical consultation. On the VAS, this effect was still present at T3: $t_{VAS(T1-T3)}(145)=2.07$, $p=0.040$, $|r|=0.17$ (figure 3). Limiting this analysis to the subgroup of patients completing the last questionnaire during days 6–13 after consultation again showed significant differences with small to medium effects for the VAS ($t_{VAS(T1-T2)}(75)=2.74$, $p=0.008$, $|r|=0.30$) and VRS ($t_{VRS(T1-T3)}(75)=2.36$, $p=0.021$, $|r|=0.26$), suggesting slight pain relief compared to T1. Surprisingly, we could not find any differences in patient-reported pain in patients undergoing a change in antirheumatic treatment (data not shown).

Pain scales in patients with RA with chronic pain and/or osteoarthritis

In patients without comorbid chronic pain or osteoarthritis, results on convergent as well as discriminant validity, reliability and responsiveness of the pain scales to changes in treatment were comparable to those of the total sample (see online supplementary file 1). Changes in pain ratings were observed for the VAS and the VRS, again suggesting decreasing pain during consultation with a longer effect on the VRS (see online supplementary file 1).

In the subgroup of patients with RA also having osteoarthritis and/or chronic pain, Spearman correlation coefficients indicated valid results with respect to convergence and discriminance (convergent validity: $r_{s,T1}=0.83-0.93$; $r_{s,T2}=0.90-0.95$; $r_{s,T3}=0.82-0.83$, $r_s^2 \geq 0.67$, all $p < 0.001$, discriminant validity: $r_{s,age}=-0.12-0.18$, $r_s^2 \leq 0.03$, $p \geq 0.34$; $r_{s,duration}=0.11-0.31$, $r_s^2 \leq 0.10$, $p \geq 0.10$). The retest-reliability coefficients were smaller, especially in view of the VAS, compared to those in the total sample, but still exceeding 0.45 ($r_{s,VAS}=0.52-0.89$, $r_{s,NRS}=0.81-0.97$, $r_{s,VRS}=0.71-0.85$, $r^2 \geq 0.27$, all $p < 0.026$). A significant change of pain ratings in patients with the aforementioned comorbidity was only observed for the VAS before and immediately after consultation ($Z_{VAS(T1-T2)}=-2.21$, $p=0.025$, $|r|=0.33$) if antirheumatic therapy was maintained. As in the total sample, this finding suggests a decrease in pain during medical consultation on the VAS.

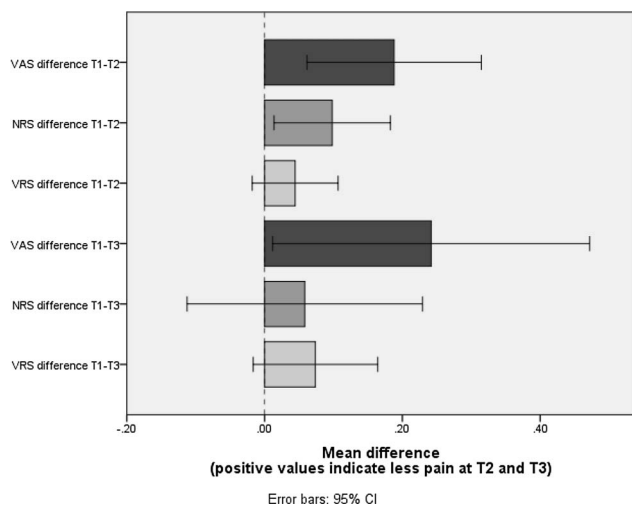


Figure 3 Mean differences of pain in relation to pain ratings before medical consultation in the total sample.

DISCUSSION

In this study, we provide a psychometric profile of the VAS, the NRS and VRS pain scales in a sample of a patient with RA derived from clinical practice. Furthermore, we took into account the potentially interfering role of physical limitations and common pain-related comorbidities in our sample. Generally speaking, construct validity and retest reliability of the pain scales can be assumed with the requirements mentioned in the literature.^{25 26} However, limitations of retest reliability of the VAS occurred when accounting for the influence of physical functioning in patients completing the last questionnaire between days 6 and 13 after medical consultation and in patients with osteoarthritis and/or chronic pain showing coefficients below the predefined criterion of $r \geq 0.7$. Overall, our results of retest-reliability correspond to previous findings.^{8–10} Apart from retest-reliability, neither physical functioning nor the presence or absence of osteoarthritis and/or chronic pain seems to considerably alter psychometric properties of single-item scales measuring overall patient-reported pain. Thus, their use in patients with RA with and without osteoarthritis and/or chronic pain, particularly in the context of outpatient clinical care, also seems feasible. For proofing the concept of discriminant validity, we favoured age and disease duration as these characteristics are supposed to be largely independent from patient-reported levels of pain. Measures of disease activity such as the DAS28 do not fulfil this requirement considering the crucial relation of pain to disease activity demonstrated in this sample.

In contrast to previous findings in the literature,^{7 11} the overall internal responsiveness of the pain scales to changes in antirheumatic treatment was low. This result was also confirmed by further subanalyses either limiting the results to patients completing the last questionnaire during days 6–13 and/or those receiving medication typically used for inflammatory pain in combination with

conventional or biological DMARDs. However, these subgroups were small, which is why the evidence level of the corresponding results is limited. We assume that this difference is most likely due to longer follow-up periods of previous studies (ranging from 2 weeks to 12 months).^{7 11–13} From the results in the literature in combination with our findings, it seems that longer follow-up periods correspond to larger SRMs, whereas responsiveness might show satisfactory results with follow-up periods of at least 2 weeks. However, in this study, we deliberately chose a rather short follow-up period in order to also cover rapid changes of pain which are especially known in patients starting or escalating glucocorticoid therapy (39.4% of patients in the total sample received glucocorticoids at the time of consultation). Moreover, in patients on stable treatment, sometimes even only a few days matter between worsening of symptoms shortly before the next scheduled intake of antirheumatic medication and rapid improvement afterwards.

Surprisingly, differences in the perception of pain before and after medical consultation were only to be found in patients maintaining antirheumatic treatment but not in patients undergoing treatment change. Although the corresponding effect size was low, this suggests decreasing pain during medical consultation in the total sample as well as in patients with and without pain-relevant comorbidity—especially when applying a comparison before and immediately after consultation on a VAS. Our results indicate that the medical consultation process itself might have a yet under-rated impact on the patient's perception of pain and that this effect might be emphasised by means of physician–patient communication. For instance, it might be hypothesised whether a rheumatologist's positive feedback signalling his satisfaction with the achieved treatment results influences his way of verbally and non-verbally communicating with a patient, which in turn improves patient-reported levels of pain. Subsequent studies will have to show whether this effect can be emphasised and to what degree it is determined by aspects of interpersonal communication or potential confounding variables such as demographic and intrapersonal factors, disease-related characteristics and the general communication setting. The corresponding pattern of results suggests, on the one hand, that significances are more likely to be found on scales with a larger degree of detail (ie, especially the VAS) while, on contrary, in some cases the denomination of all scaling points (eg, on a VRS) might also lead to less patient-reported pain compared to scales only having two anchors.

The generalisability of the results found in this patient sample seems feasible given that data are derived from a non-selected patient population with RA treated with conventional and/or biological DMARDs including all levels of disease activity and no exclusion criteria with respect to demographic characteristics. However, there are limitations with respect to this study that shall be

taken into account: first of all, although the items were obviously related to arthritic pain, we did not do a cognitive debriefing of the pain items and their anchors, which means that patients might have understood corresponding items in a manner differing from one another. Second, in order to hand out identical questionnaires to the patients, we did not include an external validation anchor/item at T2 and T3, which means that we cannot provide information on the external generalisability of our results. This information might be important for future studies implementing external validation anchors. Furthermore, we have measured responsiveness by means of SRM but not in terms of validating a change score as proposed by previous research, which might also limit our results to some extent.²⁹ Owing to the short time between the first two questionings, there might have been a recall effect which potentially contributed to the small amount of variability of pain ratings—in particular with respect to the NRS and the VRS. Furthermore, improper or missing completion date on some of the T3 questionnaires required a separate analysis with a reduced sample size for T3 questionnaires with proper completion date. However, the corresponding analyses yielded similar psychometric properties as shown above. In view of the patient sample, the number of patients receiving biological DMARD therapy was large (55.5%), which reflects a patient sample with established RA predominantly found in a specialised referral centre. Thus, results from patients with early RA might be somewhat different. Furthermore, the results of this study refer to a national sample (ie, patients from Germany with sufficient linguistic proficiency), which implies that results from a cross-cultural sample could yield somewhat different findings. Another aspect of potential limitation is the wording of the pain scales used in this study: In clinical practice and in the literature, there are various wordings and anchor denominations available for the VAS, the NRS and the VRS measuring pain—sometimes even incorporating different versions of a single scale such as the VAS.^{8 30 31} While all of these wordings might well capture patient-reported pain, we choose an identical wording and anchor denomination across all scales in this study in order to standardise patient-reported pain measurement and investigate the effect of common scaling algorithms. With the results at hand, it may be feasible to conclude that also results of items with slightly different wordings are likely to show similar findings as long as they are kept identical for all three scales and anchors as well as time frames stay the same. Additionally, cognitive mechanisms (eg, anxiety about not getting sufficient medication) or cognitive impairments might also affect pain ratings unconsciously.

In combination with available results on minimal clinically important improvement or patient acceptable symptom state of single-item pain scales in RA,³² this study completes the psychometric information profile of

these measures in patients with RA, whereas each pain scale has its individual advantages and disadvantages. In clinical trials, investigators and study sponsors may want to apply a widely accepted pain measure with a high level of detail and sound scientific background information on psychometric properties, which is frequently the reason for favouring the VAS. Its continuous scale character detects even small changes in pain perception while at the same time reducing the risk of potential recall effects associated with the use of concrete anchors on an NRS or VRS, which in turn leads to a higher variability of resulting values. Besides routine clinical practice, clinical practitioners may want to use the VAS to compare results from clinical routine to data coming from clinical trials or as part of a compound measure. However, the advantage of a VAS may also turn into the opposite with some patients having trouble in accurately reporting their level of pain on a graphical scale. For instance, the indication of pain on a VAS is supposed to be crossing the VAS line while indications beside it hinder a proper evaluation of the patient's statement. Accordingly, the VRS and the NRS are sensible alternatives in order to obtain a patient-reported pain rating. For these scales, there are various versions available with different levels of detail (ie, scaling points), whereas the appropriate level is always to be determined by the underlying purpose (eg, clinical trial, clinical routine documentation, non-interventional outcome research) and the outcome of interest. Although patients are likely to remember their previous statement on the NRS and VRS when using very short follow-up periods, their application also depends on the clinical situation and the clinicians' as well as the patients' preferences. For example, in the context of a telephone interview, it will be easier for patients to complete an NRS or a VRS instead of a VAS, while the VAS might be more helpful when addressing illiterate patients or patients lacking proficient linguistic knowledge. Although the VRS usually has less response options to choose from than the VAS or the NRS, it provides more information on each single point of the scale by using concrete statements for each response option. Accordingly, patients easily find an appropriate option according to their individual situation, which is likely to be the reason for the ongoing regular use of the VRS in routine clinical care. In conclusion, all of these three scales take about the same time to complete and are easy to use while their individual concepts are applied in a single-item or multi-item context (ie, compound measures). Nevertheless, as shown above, the use of each scale is determined by a variety of factors as well as individual preferences for or against one of these measures. The future challenge will be to investigate the potential impact of physician-patient interaction on pain in rheumatic diseases while taking into account potential confounders and individual preferences of patients and physicians in applying the pain scales at choice.

Contributors MS, EGA, GS and ME collected and analysed the data and wrote the manuscript. MS, GS and ME designed the study. MS and ME performed statistical analyses.

Competing interests None declared.

Ethics approval This study was approved by the Institutional Review Board of the University of Erlangen-Nuremberg.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med* 2011;365:2205–19.
- Gossec L, Dougados M, Rincheval N, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis* 2009;68:1680–5.
- Schmidt RF, Willis WD, eds. *Encyclopedia of pain*. Berlin: Springer, 2007.
- Englbrecht M, Turner IH, van der Heijde DM, et al. Measuring pain and efficacy of pain treatment in inflammatory arthritis: a systematic literature review. *J Rheumatol Suppl* 2012;90:3–10.
- Bradley LA. Pain measurement in arthritis. *Arthritis Care Res (Hoboken)* 1993;6:178–86.
- Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. *Pain* 2011;152:2399–404.
- Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in rheumatoid arthritis patients. *Curr Med Res Opin* 1999;15:121–7.
- Ferraz MB, Quaresma MR, Aquino LR, et al. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 1990;17:1022–4.
- Linde L, Sorensen J, Ostergaard M, et al. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. *J Rheumatol* 2008;35:1528–37.
- ten Klooster PM, Vlaar AP, Taal E, et al. The validity and reliability of the graphic rating scale and verbal rating scale for measuring pain across cultures: a study in Egyptian and Dutch women with rheumatoid arthritis. *Clin J Pain* 2006;22:827–30.
- Veehof MM, ten Klooster PM, Taal E, et al. Comparison of internal and external responsiveness of the generic Medical Outcome Study Short Form-36 (SF-36) with disease-specific measures in rheumatoid arthritis. *J Rheumatol* 2008;35:610–17.
- Verhoeven AC, Boers M, van Der Linden S. Responsiveness of the core set, response criteria, and utilities in early rheumatoid arthritis. *Ann Rheum Dis* 2000;59:966–74.
- Ward MM. Clinical measures in rheumatoid arthritis: which are most useful in assessing patients? *J Rheumatol* 1994;21:17–27.
- Clark P, Lavielle P, Martinez H. Learning from pain scales: patient perspective. *J Rheumatol* 2003;30:1584–8.
- Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scales. *Ann Rheum Dis* 1978;37:378–81.
- Herr K, Spratt KF, Garand L, et al. Evaluation of the Iowa pain thermometer and other selected pain intensity scales in younger and older adult cohorts using controlled clinical pain: a preliminary study. *Pain Med* 2007;8:585–600.
- Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011;41:1073–93.
- Butow P, Sharpe L. The impact of communication on adherence in pain management. *Pain* 2013;154(Suppl 1):S101–107.
- Edmond SN, Keefe FJ. Validating pain communication: current state of the science. *Pain* 2015;156:215–19.
- Fries JF, Spitz P, Kraines RG, et al. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
- Stucki G, Liang MH, Stucki S, et al. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum* 1995;38:795–8.
- Prevoe ML, van 't Hof MA, Kuper HH, et al. 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.
- van der Heijde DM, van 't Hof MA, van Riel PL, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;49:916–20.
- Lord FM, Novick MR. *Statistical theories of mental test scores*. Information Age Publishing Inc., 2008 (reprint).
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959;56:81–105.
- DeVon HA, Block ME, Moyle-Wright P, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007;39:155–64.
- Gossec L, Paternotte S, Aanerud GJ, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. *Ann Rheum Dis* 2011;70:935–42.
- Aletaha D, Neogi T, Silman AJ, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/ European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;69:1580–8.
- Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
- Pincus T, Bergman M, Sokka T, et al. Visual analog scales in formats other than a 10 centimeter horizontal line to assess pain and other clinical data. *J Rheumatol* 2008;35:1550–8.
- Bellamy N. *Pain measurement. Musculoskeletal clinical metrology*. Dordrecht: Kluwer Academic Publishers, 1993:65–76.
- Tubach F, Ravaud P, Martin-Mola E, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. *Arthritis Care Res (Hoboken)* 2012;64:1699–707.