



Comparison of the construct validity and reproducibility of four different types of patient-reported outcome measures (PROMs) in patients with rheumatoid arthritis

Lisanne Renskers¹ · Ron J. J. C. van Uden¹ · Anita M. P. Huis¹ · Sanne A. A. Rongen^{1,2} · Steven Teerenstra³ · Piet L. C. M. van Riel^{1,2}

Received: 27 June 2018 / Revised: 28 August 2018 / Accepted: 2 September 2018 / Published online: 12 September 2018

© The Author(s) 2018

Abstract

Patient-reported outcome measures (PROMs) are increasingly used in the assessment of patients with rheumatoid arthritis (RA). The aim of this study was to assess the construct validity and reproducibility of four types of PROMs (numerical rating scale (NRS), visual analogue scale (VAS), verbal rating scale (VRS), and Likert scale). In addition, patients' preference for one of these PROMs was measured. Patients with stable RA and no change in pain medication or steroid medication ($n = 211$) were asked to complete a questionnaire directly following, and 5 days after their outpatient visit. The questionnaire measured the validity of the PROMs in four domains (pain, fatigue, experienced disease activity, and general well-being). To assess construct validity, Pearson correlation coefficients were calculated using the Disease Activity Score-3. To assess reproducibility, intraclass correlation coefficients (ICCs) were calculated. Correlation coefficients between the different PROMs within each domain were compared. There were no differences regarding construct validity between the different PROM scale types. Regarding reproducibility, VAS and NRS scored better in the domains pain and experienced disease activity. Patients chose NRS as preferred measurement method, with the VAS in second place. The four scale types of PROMs appeared to be equally valid to assess pain, fatigue, experienced disease activity, and general well-being. VAS and NRS appeared to be more reliable for measuring pain and experienced disease activity. The patients favoured NRS. In combination with other practical advantages of NRS over VAS, we do advise the NRS as the preferred scale type for PROMs.

Keywords Outcome assessment · PROM · Psychometric properties · Rheumatoid arthritis · Self management

Introduction

Rheumatoid arthritis (RA) is an inflammatory autoimmune disease characterised by chronic joint inflammation, which leads to

pain, stiffness, function loss, and fatigue [1–3]. All of the mentioned complaints can be measured with patient-reported outcome measures (PROMs) like the visual analogue scale (VAS), the Likert scale, the verbal rating scale (VRS) and the numerical rating scale (NRS) [4–6]. During outpatient visits, PROMs are taking a more central place because they directly apply to the patients' personal perception and can identify key concerns that need to be addressed [7–10]. Studies also have shown that they correlate with the Disease Activity Score 28 (DAS28) assessed by the healthcare professional [11–14].

Recently, PROMs have become increasingly important in managing RA in daily clinical practice [6, 10, 15]. Using PROMs offers opportunities to improve understanding of patients' experiences and responses to therapy, which would otherwise be missed [7, 8, 16]. Next to this, by means of remote monitoring of disease activity with PROMs, doctors can be informed about the disease course in between visits, which could have an influence on the treatment, resulting in a

Lisanne Renskers and Ron J. J. C. van Uden contributed equally to the article.

✉ Lisanne Renskers
Lisanne.Renskers@Radboudumc.nl

¹ Radboud Institute for Health Sciences, IQ healthcare, Radboud University Medical Center, Geert Grooteplein 21, PO Box 9101 (114), 6500 HB Nijmegen, The Netherlands

² Department of Rheumatology Bernhoven, Uden, The Netherlands

³ Radboud Institute for Health Sciences, Department for Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands

better outcome in the long term. Another advantage is that PROMs can be part of patient education, leading to improvements in therapy adherence [17, 18].

There are different PROMs for measuring the different aspects of RA. Currently, the most widely used PROM scale type in rheumatology is the VAS. Using a VAS to measure patient outcomes was previously recommended in both research and daily clinical practice for its simplicity and adaptability for use in composite indices and questionnaires covering multiple domains [9, 19]. However, in more recent studies, the VAS has been more and more criticised. In many cases, its loosely defined character could undermine its construct validity and contribute to heterogeneity among patients' ratings [4, 20]. For instance, it has been found that cognition and mood influence VAS results significantly [20–22]. Furthermore, certain groups of patients have problems with using the VAS correctly [4, 9, 19, 23].

Alternatives to the VAS are the NRS, VRS and Likert scale. Especially, NRS and Likert scale are described as simple and fast to administer scales. NRS is described as easier to use than VAS, more suitable for international use and able to be used over telephone [19, 24, 25], whereas for Likert scale, research in osteoarthritis proves that Likert scales and VAS responses are highly correlated and yield similar precision, while Likert responses are easier to administer and interpret. However, for changes in Likert score, the associated VAS values can vary across a very wide range [26–28].

Studies reporting on advantages and disadvantages of the four different PROMs often report contradictory findings. Therefore, no statements can be made if one scale is generally better than the other in terms of usefulness in daily clinical practice, but rather that one scale type would be more suitable for a specific context or setting [27]. The objective of this study is to measure and compare construct validity of four scale types of PROMs within four different domains. Moreover, reproducibility of these scales and patients' preferences for PROM scale types were measured. We focused both on construct validity, defined in this study as the correlation between the PROM and the DAS28-3 score of the patient, as well as on reproducibility tested in a test-retest setting.

Materials and methods

This study was designed as a prospective longitudinal study to investigate two out of three key component criteria of the OMERACT (Outcome Measures in Rheumatology) – filter: validity and reproducibility [29] of four commonly used PROMs (NRS, VAS, VRS and Likert scale), as well as determining patients' preferred way of measuring. Adult patients with RA according to ACR/EULAR criteria [30] were included in this study. This group consisted of patients with stable disease activity, as assessed by a rheumatologist. Patients were

excluded in case they were not able to sufficiently speak and write Dutch, in case they had cognitive impairments, and/or in case the medication regime was changed during the visit.

From a population of about 900 patients with RA of Bernhoven Uden, a teaching hospital in the south of the Netherlands, each patient visiting their rheumatologist for a regular appointment at the outpatient clinic in the period from April 2016 to June 2016 was asked to participate in this research project. Flyers with information regarding this study were available in the waiting room, and each rheumatologist assessed whether patients were eligible for the study and asked them to participate. Directly after their doctor visits, patients were asked by the researcher (RvU) to complete a questionnaire on paper in which they were asked to indicate the level of pain, fatigue, experienced disease activity and general well-being they had experienced in the previous week, using NRS, VAS, VRS and Likert scale, and to indicate which PROM they preferred in order to answer the question in every domain. Next to this, all patients signed an informed consent for the use of their medical records in this research. Lastly, all patients received a self-addressed envelope containing exactly the same questionnaire to be filled out and returned by mail after 5 days. Clear instructions were given to all patients on how to fill out this questionnaire and with what purpose.

To assess construct validity, the DAS28 from the study visit was used as gold standard, but because a VAS General Health is part of the standard DAS28 score, we used a modified score. This DAS28-3 score is based on erythrocyte sedimentation rate (ESR), and a physician's assessment of tender and swollen joints of 28 joints [31]. Because the DAS28-3 score is a score of disease activity and not meant to measure pain, fatigue or general well-being, it acted for these three domains as a surrogate gold standard. The main purpose of this study was to compare the performance of the four PROM scale types within each domain, and not to determine the extent the PROMs correlate with the (surrogate) gold standard. To test the reproducibility of the aforementioned scales, patients completed the same questionnaire again 5 days after their doctor visit.

Patients were asked to describe different aspects of their disease burden (pain, fatigue, experienced disease activity and general well-being) in NRS (scaled from 1 to 10), VAS (on a 100 mm line), VRS (“Not at all, Little, Much and Very Much”) and five-point Likert scales (“Not at all, Little, Neutral, Much and Very Much”). The VRS and Likert scales were designed by the researchers and have been assessed and approved by a patient panel of the Department of Rheumatology, Bernhoven. Pain, fatigue and experienced disease activity were scored progressively (a higher score meaning more pain etcetera, or patients feeling worse), but general well-being scored the other way (a higher score meaning a patient was feeling better). A clear definition of disease activity was made, included in the questionnaire, and explained to all participants in exactly the same manner. At the end of each

specific domain, patients were asked which way of measuring they preferred by asking the question: “Which of the abovementioned methods do you prefer in order to indicate how much [pain] you experience because of your RA?”

All data were processed and analysed using SPSS version 22. Descriptive information was expressed as means and frequencies. To assess construct validity, correlation with the (surrogate) gold standard was calculated using Pearson’s correlation coefficients for each PROM, filled out for each separate domain. Although not all PROMS were continuous variables, we found it reasonable to assume that the ordered categories of Likert scales and VRS are derived from applying cutoffs to an underlying (latent) continuous scale. To check this assumption, we inspected scatterplots with 95% confidence intervals and prediction intervals to see whether correlations were indeed linear. For reproducibility intraclass correlation coefficients (ICCs) were calculated. More specifically, we tested ICC agreement. Patients who did not return their second questionnaire were excluded at this point. Pearson’s chi-square test was used to compare categorical data. For all assessments, the significance level was set at $\alpha = 0.05$. Patients’ preference was assessed with bar charts and pie charts. Lastly, we tested whether or not the PROMs’ correlation coefficients were statistically different from each other within each domain, for both construct validity [32] and reproducibility [33, 34].

The research proposal was presented to the Medical Ethical Committee of the Radboud University Medical Centre. The committee concluded that no specific obligations applied to this research.

Results

During the inclusion period, 350 RA patients visited the outpatient clinic, of whom 213 filled out the first questionnaire. The included group consists of 133 women (63%), has an average age of 63.90 years ($SD \pm 12.85$), an average DAS28-3 score of 2.99 ($SD \pm 1.09$), and a mean disease duration of 8.76 years ($SD \pm 9.06$). All patient characteristics are shown in Table 1. In Fig. 1, the inclusion flowchart is shown. When excluding ineligible patients and patients who have not returned their return envelope, 211 patients remained for the construct validity assessment and 153 for the reproducibility assessment. With regard to missing data, there was one missing value in questionnaire one. The percentage missing data in questionnaire two varied between 4 and 8%. There were no noteworthy missing data with respect to other variables. Patients with missing values did not differ from other patients with respect to age, DAS28-3 score and disease duration.

Construct validity

For each of the 211 patients included in the validity assessment, a DAS28-3 score was calculated with scores found in their patient

Table 1 Patient characteristics

Patient characteristics	<i>n</i> = 211
Female, <i>n</i> (%)	133 (63%)
Age (years), mean (SD)	63.90 (12.85)
Young, <i>n</i> (%)	49 (23.2%)
Middle, <i>n</i> (%)	88 (41.7%)
Old, <i>n</i> (%)	74 (35.1%)
DAS28-3 score, mean (\pm SD)	2.99 (\pm 1.09)
Remission, <i>n</i> (%)	71 (33.6%)
Low disease activity, <i>n</i> (%)	65 (30.8%)
Moderate disease activity, <i>n</i> (%)	67 (31.8%)
High disease activity, <i>n</i> (%)	8 (3.8%)
Disease duration (years), mean [range]	8.76 [0–46]
Short, <i>n</i> (%)	80 (37.9%)
Middle, <i>n</i> (%)	78 (37.0%)
Long, <i>n</i> (%)	53 (25.1%)

Age: young (< 55 years), middle (55–70 years), old (> 70 years). DAS28-3 Severity categories: remission (DAS28-3, < 2.6), low disease activity (DAS28-3, 2.6–3.2), moderate disease activity (DAS28-3, 3.2–5.1), high disease activity (DAS28-3, > 5.1). Disease duration: short (< 5 years), middle (5–10 years), long (> 10 years)

files. After this, Pearson correlation coefficients were calculated. These are shown in Table 2. All these correlations were non-zero except for VRS in fatigue ($p > 0.05$). In each domain, no large differences in correlation coefficients between the four PROMs were found and none of the pairwise comparisons had $p < 0.05$. The correlation coefficients for fatigue and general well-being were slightly lower than for pain and experienced disease activity. Scatterplots were made and outliers were inspected. These outliers were not different from the other patients in the study regarding age, gender, disease duration and PROM preference.

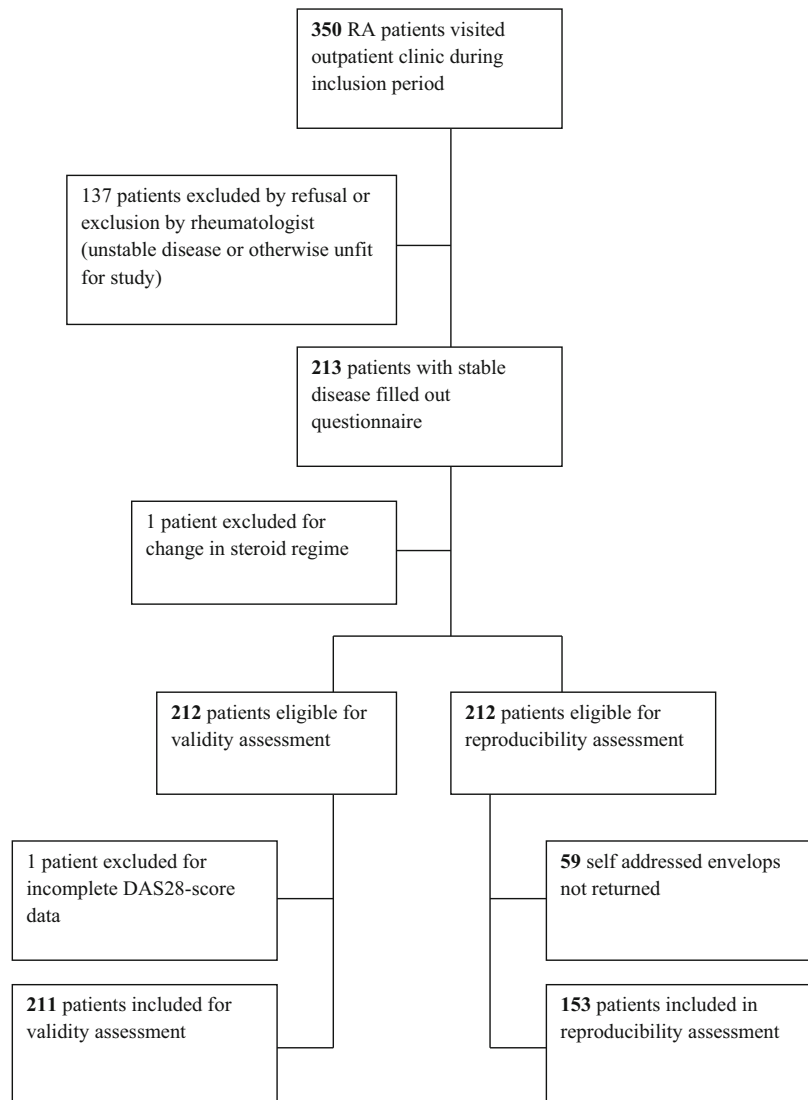
Reproducibility

Reproducibility was assessed in 153 patients with ICCs (shown in Table 3). All correlation coefficients in all scales and all domains had $p < 0.001$. When comparing the ICCs to each other within each domain, VAS and NRS appeared to be more reliable for the domains pain and experienced disease activity ($p < 0.05$) (data not shown). Regarding the domain fatigue, VAS appeared to be more reliable than Likert, while NRS was more reliable than both Likert and VRS in this domain. Lastly, there appeared to be no differences within the domain general well-being ($p > 0.05$). Moreover, there were no differences between VAS and NRS.

Patients’ preferences

When assessing patients’ preferences ($n = 212$) for PROM scale type, NRS was most preferred in every situation (varying

Fig. 1 Inclusion flowchart. Three hundred fifty patients with RA visited the outpatient clinic during the inclusion period (April 2016 to June 2016); 211 patients were included for the construct validity assessment and 153 patients for the reproducibility assessment



Inclusion flowchart. 350 patients with RA visited the outpatient clinic during the inclusion period (April 2016 to June 2016); 211 patients were included for the construct validity assessment and 153 patients for the reproducibility assessment

from 44.3 to 45.8%). VAS consistently ended in second place (20.8 to 24.5%). Third and fourth place shifted between VRS and Likert scale depending on the domain. This preference is shown in all domains, regardless of whether asked in the first or second questionnaire. There were no statistical differences between disease duration or age and PROM preference (all p values >0.05). All groups preferred NRS most. Figure 2 shows the overall preference for PROM scale type. Figure 3 shows the bar charts regarding the PROM scale type preferences per domain.

Discussion

This study aimed to compare the construct validity and reproducibility of four frequently used PROM scale types within

the domains pain, fatigue, experienced disease activity and general well-being in patients with RA. In addition, patients' preference was evaluated. True construct validity could only be assessed in the domain experienced disease activity, because the DAS28-3 score is a measure of disease activity. For the other three domains, the DAS28-3 score was used as a surrogate gold standard to compare correlations with it between the four PROMs in each domain.

All the correlation coefficients in the construct validity assessment were weak to moderate at best, as defined by Cohen [35]. The correlation coefficients were lower in fatigue than in the other domains, which can be explained by the fact that fatigue is less influenced by disease activity compared to the other constructs [36, 37], and because of its multifactorial nature [38–40]. Literature shows that there is disagreement between patients' and physicians' perspectives regarding

Table 2 Construct validity of the four PROMs in four domains ($n = 211$), correlation with the (surrogate) gold standard

Domain	PROM	<i>r</i> value	<i>p</i> value	<i>n</i> (%)
Pain	VAS	0.40	<0.001	211 (100%) ^a
	NRS	0.41	<0.001	211
	VRS	0.37	<0.001	211
	Likert	0.34	<0.001	211
Fatigue	VAS	0.16	0.02	210 (99.5%)
	NRS	0.14	0.05	211
	VRS	0.11	0.10	211
	Likert	0.14	0.05	211
Experienced disease activity	VAS	0.35	<0.001	211
	NRS	0.32	<0.001	211
	VRS	0.28	<0.001	211
	Likert	0.31	<0.001	211
General well-being	VAS	-0.25	<0.001	211
	NRS	-0.25	<0.001	211
	VRS	-0.28	<0.001	211
	Likert	-0.29	<0.001	211

Based on the results of the first questionnaire

PROM patient-reported outcome measure, VAS visual analogue scale, NRS numerical rating scale, VRS verbal rating scale; Based on results of the first questionnaire

^a Two hundred eleven patients equals a response of 100%

disease outcomes [41–44]. A study by Lati et al. found comparable correlations for VAS of patient global assessment to DAS28 score [20].

Within the construct validity assessment, the differences between the four scale types were small and appeared not different. In three out of the four domains, the NRS and VAS scales had the highest scores. VAS and NRS performed best in the domains pain and experienced disease activity. This could be explained by the fact that pain and experienced disease activity are easier to grasp concepts than a more complex subjective phenomenon as general well-being, and therefore can be scored easier using numbered values. Regarding reproducibility, VAS and NRS scored consistently better within the domains pain and experienced disease activity than VRS and Likert. Within the domain fatigue, VAS scored better than Likert and NRS scored better than both Likert and VRS.

In recent years, the VAS as tool in daily clinical practice and research has been more and more criticised. Gossec et al. described elderly, low literacy populations and some cultural groups having difficulty using the VAS, although it is a quick and easy to apply scale for pain [4]. Hjerstad et al. reported lower compliance when VAS is used compared to NRS and

VRS. They also question the ability for patients to discriminate their pain intensity beyond ten levels, and therefore also question VAS's clinical advantage over a scale such as NRS [45]. Lati et al. argued that VAS endpoints are usually not very clearly defined, and these anchors do not reflect the full range of RA activity [20].

Reproducibility of the scales in our assessment was moderate to high similar to earlier studies [9, 24, 46, 47]. The exception was the domain general well-being, all correlations both for construct validity as well as for reproducibility were lower compared to the other domains. In contrast with the other domains, the Likert and VRS scored highest in this category. Hasson et al., Bolognese et al. and Guyatt et al. described that a Likert scale is easier to use than VAS, for both clinicians as patients, although wording is important when describing complex subjective phenomena [26–28]. General well-being is such a complex subjective concept; it is quite possible that patients need an easy to use and easy to understand scale (such as Likert or VRS) to score appropriately. Neither study compared a Likert scale to NRS or VRS though.

In this study, the patients clearly preferred the NRS as scale type in all four PROM domains. The validity of this scale appeared to be at least as good as the other scales and several other studies have shown its practicality in daily clinical practice [24]. Moreover, NRS appeared to be more reliable than VRS and Likert in the domains pain, fatigue and experienced disease activity. We found no statistical differences regarding age (young, middle and old) and diagnose duration (short, middle and long) in relation to PROM preference; all groups showed a preference for NRS, regardless of domain and/or questionnaire. Hjerstad et al. also found a preference for NRS in an age-mixed population, although this was not a population purely consisting of RA patients [45]. Bellamy et al. described some key characteristics for a scale to be useful in daily clinical practice: simplicity, brevity, rapid completion and ease of scoring, and pointed out that both Likert scales and the NRS do comply to these characteristics. Next to this, the NRS is more suitable for international usage given that it is not bothered by language differences [25]. Hawker et al. described the NRS as a valid, reliable and easy and quick to administer way of measuring pain. They emphasise that the NRS can also be used verbally or by telephone, but point out the one-dimensional aspect of this scale [19]. Williamson et al. described NRS as sensitive as VAS, but much easier to use [24].

Strengths and limitations

Some nuances need to be added to our results. For reproducibility, a concern could be that high test-retest reproducibility was found due to a too short time interval between the first and second questionnaires. On this topic, Studenic et al. stated that reproducibility is dependent on the tested time interval and correlations decrease and smallest detectable differences

Table 3 Reproducibility of the four PROMs in four domains ($n = 153$)

Domain	PROM	PROM assessment 1 (mean)	PROM assessment 2 (mean)	ICC	<i>p</i> value	<i>n</i> (%)
Pain	VAS	33.51	35.52	0.85	< 0.001	144 (94.1%)
	NRS	4.16	4.34	0.82	< 0.001	144 (94.1%)
	VRS	2.16	2.23	0.68	< 0.001	143 (93.5%)
	Likert	2.51	2.58	0.67	< 0.001	146 (95.4%)
Fatigue	VAS	41.37	41.20	0.83	< 0.001	144 (94.1%)
	NRS	4.96	4.94	0.85	< 0.001	146 (95.4%)
	VRS	2.31	2.35	0.77	< 0.001	144 (94.1%)
	Likert	2.80	2.78	0.76	< 0.001	147 (96.1%)
Experienced disease activity	VAS	37.40	36.74	0.81	< 0.001	142 (92.8%)
	NRS	4.56	4.57	0.79	< 0.001	141 (92.2%)
	VRS	2.28	2.29	0.67	< 0.001	143 (93.5%)
	Likert	2.71	2.74	0.68	< 0.001	146 (95.4%)
General well-being	VAS	62.99	56.87	0.49	< 0.001	141 (92.2%)
	NRS	6.78	6.36	0.50	< 0.001	142 (92.8%)
	VRS	2.63	2.52	0.53	< 0.001	142 (92.8%)
	Likert	3.18	2.99	0.57	< 0.001	145 (94.8%)

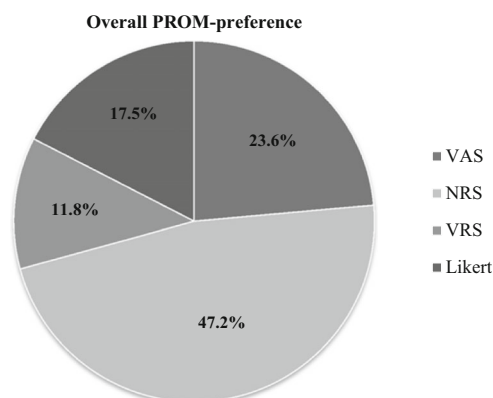
Based on complete data first and second questionnaire

PROM patient-reported outcome measure, *VAS* visual analogue scale, *NRS* numerical rating scale, *VRS* verbal rating scale, *ICC* intraclass correlation coefficient (absolute agreement)

increase with longer time intervals [48]. This can be countered by possible less response if the time interval is increased. Right now, the response rate was 72.2%, which is acceptable. When the interval is increased, one should expect this percentage to decrease, because no guarantee can be given that patients remember to return the questionnaire after a lengthened

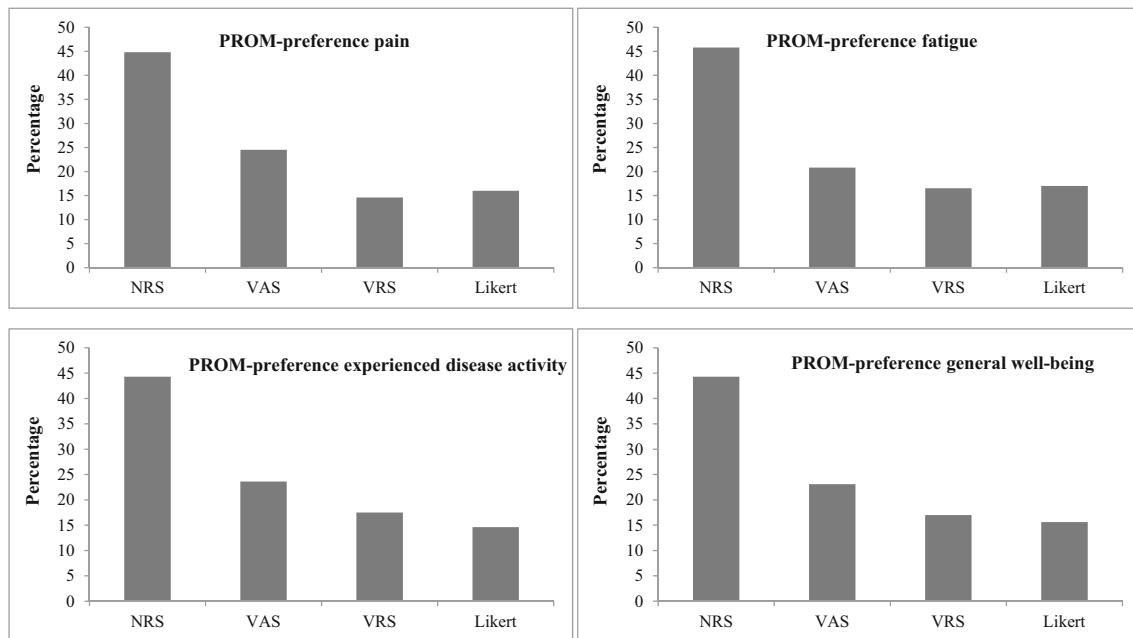
period of time. Next to this, we wanted to assess reproducibility in patients with stable disease. When increasing the time interval, we cannot be certain that all patients remain stable over this course of time. Second, the manner of measuring at baseline and after 5 days was not exactly the same, as at baseline, patients were interviewed in the hospital, while after 5 days, patients filled out the questionnaire at home. This choice was made for practical reasons: patients would otherwise have needed to visit the hospital a second time in a relative short period of time. Also, the objective of the study was to compare PROMs to each other and not to reflect the exact test-retest correlation. We expected the effect of the questionnaire being filled out at home to be equal for all PROMs and domains, and therefore, we can safely compare PROMs using our results. Next to this, for reproducibility, this would lead to a minor underestimation of correlation and not an overestimation of the results.

For preference, the vast majority for NRS preference could also be explained by the fact that NRS is a widely used measurement tool by healthcare professionals. Patients, and especially patients who were diagnosed a long time ago, could simply point out NRS as their preferred way of measuring, because they have used this for many years. However, when only assessing the preference of patients who were diagnosed with RA in the last year, this preference is still shown,



Overall PROM-preference on first questionnaire. $n = 212$ patients with Rheumatoid Arthritis. VAS: Visual Analogue Scale; NRS: Numerical Rating Scale; VRS: Verbal Rating Scale (4-point scale), Likert (5-point scale).

Fig. 2 Pie chart overall preference first questionnaire. Overall PROM preference on the first questionnaire. $n = 212$ patients with rheumatoid arthritis. VAS visual analogue scale, NRS numerical rating scale, VRS verbal rating scale (four-point scale), Likert (five-point scale)



PROM-preferences for the domains pain, fatigue, experienced disease activity, and general well-being, regarding questionnaire 1. NRS: Numerical Rating Scale; VAS: Visual Analogue Scale; VRS: Verbal Rating Scale. $n=212$ patients with Rheumatoid Arthritis

Fig. 3 PROM scale type preferences per domain for the first questionnaire. PROM preferences for the domains pain, fatigue, experienced disease activity and general well-being, regarding

questionnaire 1. NRS numerical rating scale, VAS visual analogue scale, VRS verbal rating scale. $n = 212$ patients with rheumatoid arthritis

regardless of domain of questionnaire (data not shown). Experience with NRS (or any of the other scales) due to comorbidities was not assessed.

In conclusion, this study shows that NRS is the preferred PROM scale for patients with RA and that NRS is at least as valid as the VAS, Likert or VRS. Moreover, regarding reproducibility, NRS appeared to be more reliable than VRS and Likert in order to measure pain, fatigue and experienced disease activity. Other studies have shown the practicality of the NRS over the VAS; therefore, we think that the NRS is the preferred scale type for the different PROMs.

Acknowledgements The authors thank the staff and personnel of the outpatient clinic of rheumatology in Bernhoven Uden and all patients willing to participate in this study.

Compliance with ethical standards

The research proposal was presented to the Medical Ethical Committee of the Radboud University Medical Centre. The committee concluded that no specific obligations applied to this research.

Disclosures None.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Kirwan JR, Hewlett S (2007) Patient perspective: reasons and methods for measuring fatigue in rheumatoid arthritis. *J Rheumatol* 34:1171–1173
- Carr AJ (1996) A patient-centred approach to evaluation and treatment in rheumatoid arthritis: the development of a clinical tool to measure patient-perceived handicap. *Br J Rheumatol* 35:921–932
- Dur M, Coenen M, Stoffer MA, Fialka-Moser V, Kautzky-Willer A, Kjeker I et al (2015) Do patient-reported outcome measures cover personal factors important to people with rheumatoid arthritis? A mixed methods design using the International Classification of Functioning, Disability and Health as frame of reference. *Health Qual Life Outcomes* 13(27):27. <https://doi.org/10.1186/s12955-015-0214-8>
- Gossec L (2010) Patient-reported outcomes in rheumatoid arthritis: why are they important and how should they be assessed? *Turk J Rheumatol* 25(3):99–104. <https://doi.org/10.5152/tjr.2010.11>
- Sargious A, Lee SJ (2014) Remote collection of questionnaires. *Clin Exp Rheumatol* 32(5 Suppl 85):S-168–S-172. <https://doi.org/10.5152/tjr.2010.11>
- Egsmose EL, Madsen OR (2015) Interplay between patient global assessment, pain, and fatigue and influence of other clinical disease activity measures in patients with active rheumatoid arthritis. *Clin Rheumatol* 34:1187–1194. <https://doi.org/10.1007/s11136-008-9413-7>
- Valderas JM, Alonso J, Guyatt GH (2008) Measuring patient-reported outcomes: moving from clinical trials into clinical practice. *Med J Aust* 189:93–94
- Lohr KN, Zebrack BJ (2009) Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res* 18: 99–107. <https://doi.org/10.1007/s11136-008-9413-7>
- Kalyoncu U, Dougados M, Daures JP, Gossec L (2009) Reporting of patient-reported outcomes in recent trials in rheumatoid arthritis: a systematic literature review. *Ann Rheum Dis* 68:183–190. <https://doi.org/10.1136/ard.2007.084848>

10. El Miedany Y (2014) Adopting patient-centered care in standard practice: PROMs moving toward disease-specific era. *Clin Exp Rheumatol* 32(5 Suppl 85):S40–S46
11. Pincus T, Furer V, Keystone E, Yazici Y, Bergman MJ, Luijckens K (2011) 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* 63:1142–1149. <https://doi.org/10.1002/acr.20481>
12. Gossec L, Paternotte S, Aanerud G, Balanescu A, Boumpas D, Carmona L et al (2011) 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* 70:935–942. <https://doi.org/10.1136/ard.2010.142901>
13. Leeb BF, Haindl PM, Maktari A, Nothnagl T, Rintelen B (2008) Patient-centered rheumatoid arthritis disease activity assessment by a modified RADAI. *J Rheumatol* 35:1294–1299
14. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, Bombardieri S, Choi H, Combe B, Dougados M, Emery P, Gomez-Reino J, Keystone E, Koch G, Kvien TK, Martin-Mola E, Matucci-Cerinic M, Michaud K, O'Dell J, Paulus H, Pincus T, Richards P, Simon L, Siegel J, Smolen JS, Sokka T, Strand V, Tugwell P, van der Heijde D, van Riel P, Vlad S, van Vollenhoven R, Ward M, Weinblatt M, Wells G, White B, Wolfe F, Zhang B, Zink A, Felson D, EULAR, ACR (2008) Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Arthritis Care Res* 59:1371–1377. <https://doi.org/10.1002/art.24123>
15. Walter MJ, Mohd Din SH, Hazes JM, Lesaffre E, Barendregt PJ, Luime JJ (2014) Is tightly controlled disease activity possible with online patient-reported outcomes? *J Rheumatol* 41:640–647. <https://doi.org/10.3899/jrheum.130174>
16. Wolfe F, Pincus T (1999) Listening to the patient: a practical guide to self-report questionnaires in clinical care. *Arthritis Rheum* 42:1797–1808. [https://doi.org/10.1002/1529-0131\(199909\)42:9%3C1797::AID-ANR2%3E3.0.CO;2-Q](https://doi.org/10.1002/1529-0131(199909)42:9%3C1797::AID-ANR2%3E3.0.CO;2-Q)
17. El Miedany Y (2013) PROMs in inflammatory arthritis: moving from static to dynamic. *Clin Rheumatol* 32:735–742. <https://doi.org/10.1007/s10067-013-2228-0>
18. El Miedany Y, El Gaafary M, El Arousy N, Ahmed I, Youssef S, Palmer D (2012) Arthritis education: the integration of patient-reported outcome measures and patient self-management. *Clin Exp Rheumatol* 30:899–904
19. Hawker GA, Mian S, Kendzerska T, French M (2011) Measures of adult pain: visual analog scale for pain (VAS pain), numeric rating scale for pain (NRS pain), McGill pain questionnaire (MPQ), short-form McGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP). *Arthritis Care Res* 63(Suppl 11):S240–S252. <https://doi.org/10.1002/acr.20543>
20. Lati C, Guthrie LC, Ward MM (2010) Comparison of the construct validity and sensitivity to change of the visual analog scale and a modified rating scale as measures of patient global assessment in rheumatoid arthritis. *J Rheumatol* 37:717–722. <https://doi.org/10.3899/jrheum.090764>
21. Luna KS, Ortiz AM, Patino E, Aguilera C, Velasco T, Garcia de Vicuna R et al (2012) Influence of the structure of mood in the assessment of rheumatoid arthritis through the visual analog scale for pain, HAQ and DAS28. *Reumatol Clin* 8:328–333. <https://doi.org/10.1016/j.reuma.2012.04.002>
22. Cordingley L, Prajapati R, Plant D, Maskell D, Morgan C, Ali FR, Morgan AW, Wilson AG, Isaacs JD, the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS), Barton A (2014) Impact of psychological factors on subjective disease activity assessments in patients with severe rheumatoid arthritis. *Arthritis Care Res* 66:861–868. <https://doi.org/10.1002/acr.22249>
23. Nicolau G, Yogui MM, Vallochi TL, Gianini RJ, Laurindo IM, Novaes GS (2004) Sources of discrepancy in patient and physician global assessments of rheumatoid arthritis disease activity. *J Rheumatol* 31:1293–1296
24. Williamson A, Hoggart B (2005) Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 14:798–804. <https://doi.org/10.1111/j.1365-2702.2005.01121.x>
25. Bellamy N, Campbell J, Syrotuik J (1999) Comparative study of self-rating pain scales in osteoarthritis patients. *Curr Med Res Opin* 15:113–119. <https://doi.org/10.1185/03007999909113372>
26. Guyatt GH, Townsend M, Berhan LB, Keller JL (1987) A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis* 40:1129–1133
27. Hasson D, Arnetz BB (2005) Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. *Int Electron J Health Educ* 8:178–192
28. Bolognese JA, Schnitzer TJ, Ehrich EW (2003) Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthr Cartil* 11:499–507. [https://doi.org/10.1016/S1063-4584\(03\)00082-7](https://doi.org/10.1016/S1063-4584(03)00082-7)
29. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, Conaghan PG, Bingham CO III, Brooks P, Landewé R, March L, Simon LS, Singh JA, Strand V, Tugwell P (2014) Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 67:745–753. <https://doi.org/10.1016/j.jclinepi.2013.11.013>
30. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO et al (2010) 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheumatol* 62:2569–2581. <https://doi.org/10.1002/art.27584>
31. Prevoe ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL (1995) 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* 38:44–48. <https://doi.org/10.1002/art.1780380107>
32. Lee IA, Preacher KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [Comput Softw]. September 2013. Available at: <http://quantpsy.org>
33. Silver NC, Hittner JB, May K (2004) Testing dependent correlations with nonoverlapping variables: a Monte Carlo simulation. *J Exp Educ* 73:53–69
34. Diedenhofen B, Musch J (2015) cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 10:e0121945. <https://doi.org/10.1371/journal.pone.0121945>
35. Cohen J (1998) *Statistical power analysis for the behavioral sciences*, 2nd edn. Erlbaum, Hillsdale
36. Pollard L, Choy E, Gonzalez J, Khoshaba B, Scott D (2006) Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology (Oxford)* 45:885–889. <https://doi.org/10.1093/rheumatology/kei021>
37. Jump RL, Robinson ME, Armstrong AE, Barnes EV, Kilbourn KM, Richards HB (2005) Fatigue in systemic lupus erythematosus: contributions of disease activity, pain, depression, and perceived social support. *J Rheumatol* 32:1699–1705
38. Tack BB (1990) Fatigue in rheumatoid arthritis: conditions, strategies, and consequences. *Arthritis Rheum* 3:65–70. <https://doi.org/10.1002/art.1790030203>
39. Hewlett S, Chalder T, Choy E, Cramp F, Davis B, Dures E, Nicholls C, Kirwan J (2010) Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology (Oxford)* 50:1004–1006. <https://doi.org/10.1093/rheumatology/keq282>
40. Hewlett S, Nicklin J, Treharne GJ (2008) Fatigue in musculoskeletal conditions. *Topical Reviews: Reports on the Rheumatic*

- Diseases Series 6:1 <https://www.arthritisresearchuk.org/health-professionals-and-students/reports/topical-reviews/topical-reviews-autumn-2008.aspx>
41. Leeb B, Sautner J, Leeb B, Fassel C, Rintelen B (2006) Lack of agreement between patients' and physicians' perspectives of rheumatoid arthritis disease activity changes. *Scand J Rheumatol* 35: 441–446. <https://doi.org/10.1080/03009740600906727>
 42. Hewlett SA (2003) Patients and clinicians have different perspectives on outcomes in arthritis. *J Rheumatol* 30:877–879
 43. Suarez-Almazor ME, Conner-Spady B, Kendall CJ, Russell AS, Skeith K (2001) Lack of congruence in the ratings of patients' health status by patients and their physicians. *Med Decis Mak* 21: 113–121. <https://doi.org/10.1177/0272989X0102100204>
 44. Taylor PC (2010) The importance of the patients' experience of RA compared with clinical measures of disease activity. *Clin Exp Rheumatol* 28(3 Suppl 59):S28–S31
 45. Hjerstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N, Kaasa S, European Palliative Care Research Collaborative (EPCRC) (2011) 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 Manag* 41:1073–1093. <https://doi.org/10.1016/j.jpainsymman.2010.08.016>
 46. Uhlig T, Kvien TK, Pincus T (2009) Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. *Ann Rheum Dis* 68:972–975. <https://doi.org/10.1136/ard.2008.097345>
 47. Clark P, Lavielle P, Martinez H (2003) Learning from pain scales: patient perspective. *J Rheumatol* 30:1584–1588
 48. Studenic P, Stamm T, Smolen JS, Aletaha D (2016) Reliability of patient-reported outcomes in rheumatoid arthritis patients: an observational prospective study. *Rheumatol* 55:41–48. <https://doi.org/10.1093/rheumatology/kev282>