Journal of Pain Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Test–Retest Reliability of the Generalized Pain Questionnaire in Patients with Rheumatoid Arthritis and Preliminary Reference Values for Non-Clinical and Several Clinical Samples

Niels Jansen[1], Peter M ten Klooster[2], Harald E Vonkeman[3], Jan R Buitenweg[1]

[1]Biomedical Signals and Systems, University of Twente, Enschede, the Netherlands; [2]Psychology, Health & Technology, University of Twente, Enschede, the Netherlands; [3]Department of Rheumatology and Clinical Immunology, Medisch Spectrum Twente, Enschede, the Netherlands

Correspondence: Niels Jansen, Email n.jansen@utwente.nl

Introduction: Generalized pain hypersensitivity is a characteristic feature in many different types of chronic pain. Recently, a 7-item self-reported Generalized Pain Questionnaire (GPQ) was developed to evaluate the presence and severity of generalized pain hypersensitivity in chronic pain patients. Here, we evaluate the test–retest reliability of the GPQ and report on preliminary reference values for various patient groups and healthy subjects.

Methods: Eighty-five patients diagnosed with Rheumatoid Arthritis (RA) completed the GPQ twice over a 2-week interval. Relative and absolute indicators of reliability were determined using data of 69 patients (81.2% retest response rate). Using readily available datasets, preliminary reference data were established in two nonclinical populations (NCP1; N = 30 and NCP2; N = 111), and for patients diagnosed with RA (N = 114), gout (N = 97), fibromyalgia (N=98), or neuropathy (N = 25), or participants in a pain rehabilitation program (N = 33).

Results: Total GPQ scores had an ICC of 0.78 (95% CI: 0.67 to 0.86). While no systematic or proportional differences were found for the GPQ total score; two (near-)significant systematic differences were observed for the individual questions. The standard error of measurement and minimal detectable change were 2.22 and 6.2, respectively. Mean ± SD scores were found to be 0.8 ± 1.2 (NCP1), 4.0 ± 4.6 (NCP2), 6.4 ± 5.5 (Gout), 6.5 ± 5.1 (RA), 8.1 ± 4.5 (Neuropathy), 13.6 ± 4.0 (Rehabilitation) and 16.0 ± 5.0 (Fibromyalgia).

Discussion: This study shows that the GPQ has acceptable reliability to be used as a tool to evaluate the presence and intensity of generalized pain hypersensitivity. The absolute measures of reliability and the preliminary reference values reported here aid in the interpretation of future studies with the GPQ.

Keywords: generalized pain hypersensitivity, widespread pain, preliminary reference values, reliability, face validity

Introduction

A characteristic feature of several different types of chronic pain syndromes is the presence of widespread hypersensitivity to pain, also called generalized pain hypersensitivity. The mechanism underlying this phenomenon is thought to be central sensitization, ie increased responsiveness of nociceptive neurons in the central nervous system to their normal or subthreshold afferent input.[1] Typically, the presence of generalized pain hypersensitivity is probed by performing pain threshold measurements at a local (ie painful) site of the body, and at one remote body site.[2] Upon comparing the findings with a reference database, finding lower pain thresholds not only at the local site but also at the remote body site is an indication of generalized pain hypersensitivity. In this way, the presence of generalized pain hypersensitivity has been observed in various chronic pain syndromes.[3]

In many research settings, but even more so in a clinical setting, the evaluation of generalized pain hypersensitivity using pain threshold measurements can be impractical as this is time-consuming, complex and requires specialized equipment.[4] For that reason, the Generalized Pain Questionnaire (GPQ) was recently developed as a time-efficient alternative.[5] The GPQ is

Received: 18 July 2023
Accepted: 3 November 2023
Published: 4 December 2023

a 7-item self-report questionnaire which aims to screen for the presence and intensity of typical generalized pain manifestations. The GPQ demonstrated high internal consistency and reliability and could accurately distinguish patients with rheumatoid arthritis from patients with fibromyalgia[5] – a condition in which widespread chronic pain is thought to be caused by central maladaptive mechanisms.[6,7] GPQ scores additionally showed the expected pattern of associations with the other patient-reported measures, including pain intensity, neuropathic-like pain features, physical disability and health-related quality of life, supporting its construct validity.[5]

More recently, a study has been performed in which the convergent validity of the GPQ against quantitative sensory testing (QST) was explored in rheumatoid arthritis (RA) patients.[8] This study also included a subgroup of patients in which central maladaptive mechanisms were suspected to predominantly underlie the pain. In this study, electrical- and pressure pain threshold measurements were performed at different body locations. On most of these body locations, a low but significant negative correlation between the GPQ and the pain threshold measurements was found in the total group ($r \approx -0.35$), which increased into a moderate negative correlation in the subgroup in which central maladaptive mechanisms were suspected ($r \approx -0.5$).[8] These findings provide preliminary evidence for the convergent validity of the self-reported GPQ with experimental pain threshold measurements.

Taken together, the earlier studies[5,8] performed with the GPQ indicate that the questionnaire has both face- and convergent validity. To date, however, no information is available with regard to the test–retest reliability of the questionnaire. This information is useful as it provides insights into the extent of random temporal fluctuations in GPQ scores.[9] This information can help to improve the GPQ, aid the future design of studies, and assist in the interpretation of changes in the GPQ scores over time (ie distinguishing random score differences from true changes). Moreover, no (preliminary) reference values have been provided yet for the GPQ in different patient populations. First, such reference values provide further information on the face validity of the GPQ in non-clinical or pain-free samples versus clinical populations where central mechanisms are assumed to be involved to different extents. Second, these preliminary reference values can help in the interpretation (ie whether generalized pain hypersensitivity is present and in what intensity) of GPQ scores in both research and clinical settings.

Therefore, in this study, the test–retest reliability of the GPQ is evaluated in a group of RA patients. Second, preliminary reference values of different groups of (non-)clinical subjects are reported upon. In addition to two datasets including nonclinical participants, reference values will be shown for patients participating in a pain rehabilitation program, as well as of patients diagnosed with RA, fibromyalgia, gout or neuropathy.

Methods

Test–Retest Reliability: Participants

From September 2020 until July 2022, patients diagnosed with Rheumatoid Arthritis (RA) were recruited for participation in this study. In total 85 patients were included in the study. Sixteen patients did not return the second GPQ questionnaire (retest), resulting in a dataset of 69 patients (81.2% response rate) for the test–retest analysis. Of these 69 patients, in total 38 (55%) were female, with the mean age being 62.1 (SD: 11.5). The average disease duration of the patients was 14.9 (SD: 9.0). The data for this study were mostly collected as part of a larger study on the assessment of generalized pain hypersensitivity in RA.[8] This study protocol was approved by a medical ethical committee (MEC-U, reference number: NL73282.100.20), and the study was conducted in accordance with the 1964 Helsinki Declaration and its later amendments. Furthermore, the study was preregistered in the Netherlands Trial Register with trial ID NL8760. All patients were sent the written information regarding the study prior to participation, and all patients signed informed consent.

Of the analyzable dataset (N = 69), for the first 39 RA-patients the first GPQ measurement was combined with additional data collection during an outpatient clinic visit. For this subgroup of 39 patients, the first time the GPQ was filled in was at the end of a more comprehensive, 1-hour measurement session conducted at the Medisch Spectrum Twente (MST; Enschede, The Netherlands). The GPQ was filled in on paper and subsequently checked by the researcher for completeness. In the case the questionnaire contained missing values, the patient was asked to fill in the remaining items. During this measurement session, the GPQ was filled in together with other questionnaires (Central Sensitization

Inventory [CSI] and Pain Catastrophizing Scale [PCS]), and after patients underwent three separate QST measurements: electrical detection- and pain threshold measurements and pressure pain threshold measurements.[8] At the end of the measurement session, the patient was given another GPQ form with return envelope and asked to fill in the questionnaire at home, two weeks after the visit. For safety reasons and to obtain valid measurements from the QST measurements, patients were excluded from participation in case the patient had an implanted stimulation device, was diagnosed with diabetes or psoriatic arthritis, in case of pregnancy, or when language barriers were present. All these exclusion criteria were evaluated by asking the patients for these conditions.

To obtain the a-priori determined sample size of at least 61 patients for a sufficiently precise estimate of the test–retest reliability, we continued to recruit patients. For these patients recruited at a later stage (N = 30 in total), both the first and the second questionnaire were filled in at home and returned, again with a two-week interval. For these patients, no exclusion criteria were set as no QST measurements were performed.

Preliminary Reference Values: Datasets

For the preliminary reference values, multiple readily available datasets from different convenience sample studies were used. In two studies, non-clinical participants were included. One important distinction between these two non-clinical populations is that in sample 1 (NCP1) one of the exclusion criteria was that participants had to be pain free at the moment of filling in the GPQ, while in the other study (NCP2) no exclusion criteria were set. In five studies, different patient groups were included. Three of these included patients with clear clinical diagnoses (gout, rheumatoid arthritis (RA) and fibromyalgia (FM)). The remaining two included more heterogenous patient groups (patients participating in a pain rehabilitation program (RHB) and patients with neuropathy). In all but the NCP1 dataset, also the responses to the PainDETECT questionnaire[10] were collected. From this questionnaire, the numerical rating score ranging from 0 to 10 at the moment of filling in the questionnaire ("NRS Now") could be obtained, as well as the strongest ("NRS Worst") and average pain ("NRS Average") during the 4 weeks prior to filling in the questionnaire. Below in-text and in Table 1, per dataset relevant demographics and information are provided.

- Non-clinical population 1 (NCP1). In this dataset, a sample of primarily (but not exclusively) employees (N = 30) of the Roessingh Rehabilitation Centre (RRC) in Enschede, The Netherlands, were recruited to participate in a larger study. Data of the GPQ was obtained as part of a larger study, in which prior to filling in the questionnaire, (electrical) pain threshold measurements were performed. Measurements took place between April 2018 and June 2019. The GPQ was filled in on paper and was subsequently checked for completeness by a researcher. Participants were excluded from

Table 1 Demographic information of the seven studies included for the determination of preliminary reference values

Subjects	DS Abbr.	N	Age (M, SD)	Females (N, %)	NRS Now (M, SD)	NRS Worst (M, SD)	NRS Average (M, SD)	Reference to Paper
Nonclinical pop. #1	NCP1	30	49 (9)	16 (53)	1.1 (0.6)*	NA	NA	NA; To be submitted
Nonclinical pop. #2	NCP2	111	45 (13)	90 (81)	1.8 (2.3)	4.1 (3.0)	2.5 (2.5)	NA
Gout	Gout	92	68 (12)	80 (87)	2.0 (2.4)	3.1 (3.4)	2.3 (2.7)	[11]
Rheumatoid Arthritis	RA	114	60 (12)	76 (67)	3.1 (2.6)	4.5 (3.0)	3.6 (2.6)	[5]
Neuropathy	NP	25	65 (7)	9 (36)	4.8 (2.1)	6.7 (1.9)	5.1 (1.9)	NA
Rehabilitation	RHB	33	47 (11)	19 (58)	5.9 (2.2)	8.3 (1.2)	6.6 (1.7)	NA; To be submitted
Fibromyalgia	FM	98	45 (12)	88 (90)	6.7 (1.7)	8.1 (1.3)	6.9 (1.5)	[5]

Notes: *In the NCP1 dataset, the NRS Now has a scale between 1 ("no pain") and 10 ("worst pain imaginable"). In all other datasets, the scale ranges from 0 ("no pain") to 10 ("worst pain imaginable").
Abbreviations: DS, Dataset; Abbr, Abbreviation; NRS, Numerical Rating Scale; M, Mean; SD, Standard Deviation; NA, Not Applicable.

participation in the case any condition resulting in chronic pain was present. Further exclusion criteria were the presence of language problems, or in the case the participant had diabetes, an implanted stimulation device or in case of pregnancy. For more information on the characteristics of the participants, see Table 1.

- Non-clinical population 2 (NCP2). In this dataset collected in 2021, a convenience sample of employees (N = 111) of the Medisch Spectrum Twente (MST) hospital in Enschede, The Netherlands, was asked to fill in the GPQ online via Qualtrics. Also several other questions were asked, amongst others if the participant had pain regularly and a numerical rating scale (NRS) for average pain in the past 4 weeks. In total 48 (43%) responded experiencing regular pain and 34 (31%) reported an average NRS pain score ≥4. No exclusion criteria were set, eg, on the presence of (former) (chronic) pain. For more information on the characteristics of the participants, see Table 1.
- Gout. In this dataset, patients (N = 92) were recruited from the rheumatology outpatient department of the MST in Enschede, The Netherlands. The questionnaire was filled in on paper at home and sent back to the MST for further analysis in 2019. All patients were diagnosed with crystal proven gout as observed by monosodium urate crystals in the synovial fluid using a polarized light microscope. The diagnosis was established on or before 2018. No further inclusion or exclusion criteria were set. For more specific information on the characteristics of the patients and the procedures followed to obtain the dataset, see Ten Klooster et al[11] and Table 1.
- Rheumatoid Arthritis (RA). In this dataset, 114 patients (N = 114) from the department of the MST in Enschede, The Netherlands completed the GPQ within the Dutch Rheumatoid Arthritis Monitoring (DREAM) RA registry. In this quality registry system, patient-reported and clinical data from patients diagnosed with Rheumatoid Arthritis are prospectively registered. No further inclusion or exclusion criteria were set. Measurements were completed online in 2017. For more specific information on the characteristics of the patients and on the procedures followed to obtain the dataset, see van Bemmel et al[5] and Table 1.
- Neuropathy (NP). In this dataset, patients (N = 25) were included at the MST in Enschede, The Netherlands. All patients had been diagnosed by a neurosurgeon with neuropathic pain. The questionnaires were filled in online in 2021–2022. No further inclusion or exclusion criteria were set. For more specific information on the characteristics of the patients, see Table 1.
- Rehabilitation (RHB). In this dataset, patients (N = 33) were included which participated in an inpatient pain rehabilitation program at the RRC, in Enschede, The Netherlands. In the pain rehabilitation program, the participants follow a 10-week program based on Acceptance and Commitment Therapy.[12] Further information, eg, on exclusion criteria or setting wherein the GPQ was filled in, has already provided with the description of the NCP1 dataset. The majority (N = 19) of this group was diagnosed with "generalized pain" (ICD10 52.9). Other diagnoses were related to regional musculoskeletal pain (N = 7), fibromyalgia (N = 2), and others (N = 5). For more information on the characteristics of the patients, see Table 1.
- Fibromyalgia (FM). In this dataset, patients (N = 98) with an ICD10 code M79.7 were recruited from the rheumatology outpatient department of the MST in Enschede, The Netherlands. No further inclusion or exclusion criteria were set. The questionnaire was filled in on paper and sent back to the MST for further analysis. Measurements took place in 2017. For more specific information on the characteristics of the patients and on the procedures followed to obtain the dataset, see van Bemmel et al[5] and Table 1.

Generalized Pain Questionnaire

The generalized pain questionnaire (GPQ) is a 7-item self-report instrument. In this study, the Dutch version of the GPQ was used. The instruction of the GPQ asks subjects to rate their general degree of complaints (without a recall period) on 7 possible pain experiences on a 5-point scale from o ("never") to 4 ("very strongly"). The items on the GPQ assess the presence and intensity of various symptoms commonly associated with patients with likely generalized pain hypersensitivity.[5,11] Total scores can range from 0 to 28, whereby a cutoff value of ≥11 is proposed as an indicator of likely generalized pain hypersensitivity.[5] Cronbach's alpha was found to be good in the current study with both the test ($\alpha = 0.82$) and the retest ($\alpha = 0.87$) administration. Except for instructions to fill in the provided questionnaire, no further instructions were provided.

Statistical Analyses

The relative test–retest reliability of total GPQ scores and individual item scores was evaluated by calculating intraclass correlation coefficients (ICCs) using a two-way mixed effects model with absolute agreements.[13,14] Following the recommendations of Portney,[15] an ICC is considered excellent in case the computed estimate is higher than 0.9. Between 0.75 and 0.9, the ICC is considered good, and between 0.5–0.75 and below 0.5, the ICC is considered moderate and poor, respectively. The test–retest reliability of the criterion of ≥11 on the total GPQ scores as an indicator for likely generalized pain hypersensitivity[5] was tested by computing Cohen's Kappa (unweighted). Following Landis and Koch[16] interpretation of the outcome, the agreement is considered excellent above 0.8, substantial between 0.6 and 0.8, moderate between 0.4 and 0.6 and poor below 0.4. A Bland-Altman plot was constructed[17] to graphically assess the variability between the repeated measurements and to screen for any systematic or proportional biases. To statistically test for the presence of a systematic bias, a dependent-samples *t*-test was performed. To statistically test for the presence of a proportional bias, a linear regression model was fitted on the data. Statistical significance was considered when $p < 0.05$. Results are provided as mean including the standard deviation except if stated otherwise. MATLAB (2019b, MathWorks, Inc) was used for statistical testing and visualization of the results.

As absolute reliability measures, the standard error of measurement (SEM) and the minimal detectable change (MDC) on an individual were calculated. The SEM is an indication of the amount of variation between tests,[18] and is calculated by dividing the standard deviation of the within-subject differences between test–retest scores (test score – retest score) by $\sqrt{2}$.[9] Subsequently, the MDC – which provides an indication of the change in score between measurements beyond measurement error – at 95% confidence level can be calculated by: $1.96 * \sqrt{2} * SEM$.[19]

For the preliminary reference values, data from a subject was only considered in case maximally one item was missing. Only in the gout dataset, this resulted in the exclusion of the data from in total 5 subjects. In all other datasets, no or only one item was missing per subject. If one item was missing, the total score was computed as the sum score of the remaining six items.

Results

Test–Retest Reliability

Out of the 85 participants who filled in the first questionnaire, 69 also filled in the second questionnaire (81.2% response rate). On average, the second questionnaire was filled in 15.0 (SD: 3.2; min = 10 days; max = 27 days) days after the first questionnaire. Of the included participants, at both the first and the second measurement 1 participant had 1 missing item response (Q4 and Q5, respectively; not the same subject).

The ICC of the total score was found to be 0.78 with a lower and upper bound (95% CI) of 0.67 and 0.86. The mean total scores of the GPQ were 6.9 (SD: 4.8) on the first measurement and 7.6 (SD: 4.9) on the second measurement. Cohen's kappa for the agreement between meeting or not meeting the likely generalized pain hypersensitivity criterion at the test- and retest administration was found to be substantial at 0.66 with a lower and upper bound (95% CI) of 0.45 and 0.87. Between the first and second measurement, no systematic differences were found ($p = 0.09$). The Bland-Altman plot (Figure 1) additionally showed that differences in test–retest scores were reasonably evenly distributed across and not linearly associated with their average values, indicating no systematic proportional bias. The SEM for total GPQ scores was found to be 2.22 and the MDC 0.74, respectively.

Analysis per individual item score revealed that ICCs ranged from 0.41 (Q2) to 0.72 (Q7). To see the ICC per individual item and for the associated confidence intervals, see Table 2. No proportional errors were observed in any of the individual items. With the third question (Q3), a statistically significant higher score was observed at the retest as compared to the test ($p < 0.001$).

Preliminary Reference Values

In the non-clinical populations, the lowest GPQ total scores were found, with mean (SD) scores of 0.8 (1.2) and 4.0 (4.6), respectively, for the NCP1 and NCP2 dataset. For the datasets with gout, RA and neuropathy patients, the mean scores and standard deviations were found to be, respectively, 6.4 (5.5), 6.5 (5.1) and 8.1 (4.5). The highest scores were found in

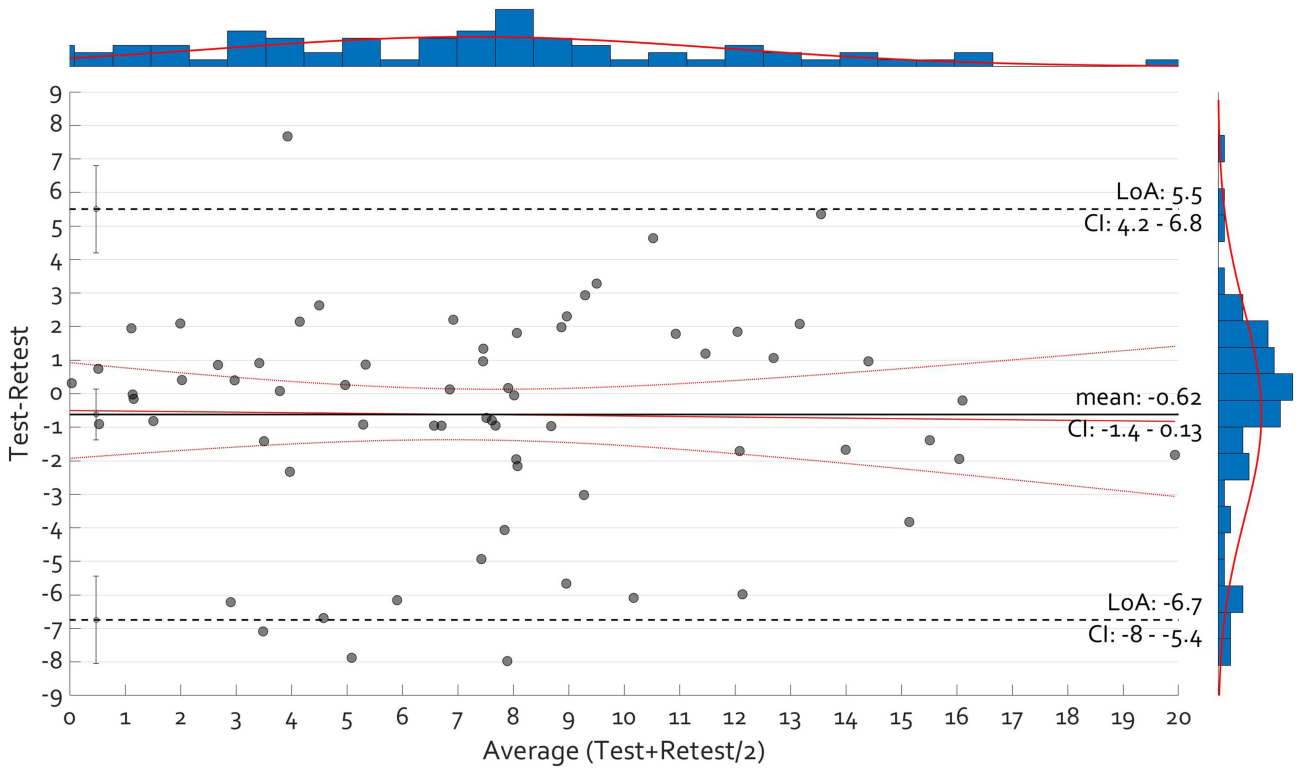

Figure 1 Bland-Altman plot of the GPQ total score. To prevent identical scores from overlapping, jitter has been added to all datapoints. On the top and on the far right, histograms providing information on the distribution of the data is added. Moreover, in these plots the red line indicates a normality fit on the data.
Abbreviations: LoA, Limit of Agreement; CI, Confidence Interval (95%).

the rehabilitation and FM dataset, with mean scores and standard deviations of, respectively, 13.6 (4.0) and 16.0 (5.0). For additional descriptive statistics such as the median, interquartile range and the minimum and maximum value per dataset, see Table 3.

Correspondingly, the prevalence of likely generalized pain hypersensitivity (ie a GPQ total score ≥11; see van Bemmel et al[5]) was found to be highest in the Rehabilitation (82%) and FM (80%) datasets, followed by the datasets which included patients with Neuropathy (32%), RA (23%) and Gout (19%). The lowest prevalences of likely generalized pain hypersensitivity was found in the NCP1 (0%) dataset followed by the NCBP2 (11%) dataset. For a visualization of the GPQ total score distributions per dataset and the prevalences of likely generalized pain hypersensitivity, see Figure 2.

Table 2 Individual item evaluations. Per individual item of the GPQ, the table provides information on the average score on the test and retest. Moreover, the computed ICC (including 95% confidence interval) is shown, as well as whether a proportional or systematic error has been found

	Item	Mean (SD), M1	Mean (SD), M2	ICC	System. Error (p)	Prop. Error (p)	Abs. Agr. (%)
Q1	Pain from light touch (eg from a pat on the back or handshake)	2.0 (1.0)	1.9 (0.9)	0.64 [0.47–0.76]	0.45	0.16	55
Q2	Pain from friction on skin (eg from clothing or the wind)	1.4 (0.6)	1.5 (0.6)	0.41 [0.20–0.59]	0.07	0.33	64
Q3	Pain from heat or cold that most people would not experience as painful (eg from cold water or holding cold objects)	1.7 (1.0)	2.1 (1.0)	0.65 [0.44–0.78]	<0.001	0.68	57
Q4	Pain that lasts longer than with most other people	2.4 (1.1)	2.6 (1.0)	0.59 [0.41–0.73]	0.10	0.54	49

(*Continued*)

Table 2 (Continued).

	Item	Mean (SD), M1	Mean (SD), M2	ICC	System. Error (p)	Prop. Error (p)	Abs. Agr. (%)
Q5	Paint hat arises only later and that would not arise in most other people (eg hours later or the next day after exertion, such as walking)	2.6 (1.1)	2.8 (1.0)	0.60 [0.42–0.73]	0.36	0.28	57
Q6	Unusually intense experiences of pain (eg nausea or gasping for air)	1.6 (0.9)	1.5 (0.8)	0.64 [0.48–0.76]	0.63	0.10	72
Q7	Pain that also spreads to other parts of the body (eg pain in the hand that spreads to the underarm when holding objects)	2.3 (1.1)	2.3 (1.1)	0.72 [0.58–0.82]	0.67	0.63	54

Abbreviations: Q, Question (eg Q1 = Question 1); SD, Standard Deviation; M1, Measurement 1; M2, Measurement 2; ICC, Intraclass Correlation Coefficient; System. Error, Systematic Error; Prop. Error, Proportional Error; Abs. Agr, Absolute Agreement.

Table 3 Preliminary reference values per dataset.

Dataset (Description)	GPQ >10 (%)	Mean	Standard Deviation	Median	IQR	Minimum	Maximum
NCP1	0	0.8	1.2	0	1	0	5
NCP2	11	4.0	4.6	2	6.8	0	18
Gout	20	6.4	5.5	6	10	0	23
RA	23	6.5	5.1	6	8	0	22
Neuropathy	32	8.1	4.5	8	6.5	0	18
RHB	82	13.6	4.0	14.5	5	4	22
FM	80	16.0	5.0	16	7.3	6	26

Notes: For All the 7 datasets, the mean, standard deviation, median, interquartile range (IQR) and the minimum and maximum of the GPQ total score are provided. Also the percentage of subjects in a dataset which scored ≥11 is provided, which is the cut-off value suggested by van Bemmel et al[5] to identify possible generalized pain hypersensitivity.
Abbreviations: NCP1, Nonclinical Population; RA, Rheumatoid Arthritis; RHB, Rehabilitation; FM, Fibromyalgia; IQR, Interquartile Range.

Discussion

Earlier research with the GPQ[5,8] indicated that the questionnaire seems to hold promise for evaluating the presence and intensity of generalized hypersensitivity to pain. Thus far, no information was, however, available regarding the test–retest reliability of the questionnaire nor was information available on reference values. In this study, we have addressed both gaps in knowledge by evaluating the two-week test–retest reliability of the GPQ questionnaire and by providing preliminary reference values in several non-clinical and clinical populations.

Test–Retest Reliability

For the GPQ total score, we found an ICC of 0.78 with a 95% CI ranging from 0.67 to 0.86. No systematic or proportional bias was identified in the Bland-Altman analysis. The relative reliability estimate of the GPQ is comparable to those that have been found, for instance, for other short screeners like the pain sensitivity questionnaire[4] and the PainDETECT questionnaire,[20] but somewhat lower as compared to the longer 25-item central sensitivity inventory (CSI).[21] With also taking into account the computed confidence interval of the ICC,[13] the reliability of the total GPQ score could be considered "moderate" to "good".[15] Moreover, substantial agreement was found between the two measurements when using a cut-off value of ≥11 for the evaluation of likely generalized pain hypersensitivity. In view of the envisioned use of the GPQ as a tool to be used within research settings or as a screener – and not as a diagnostic –

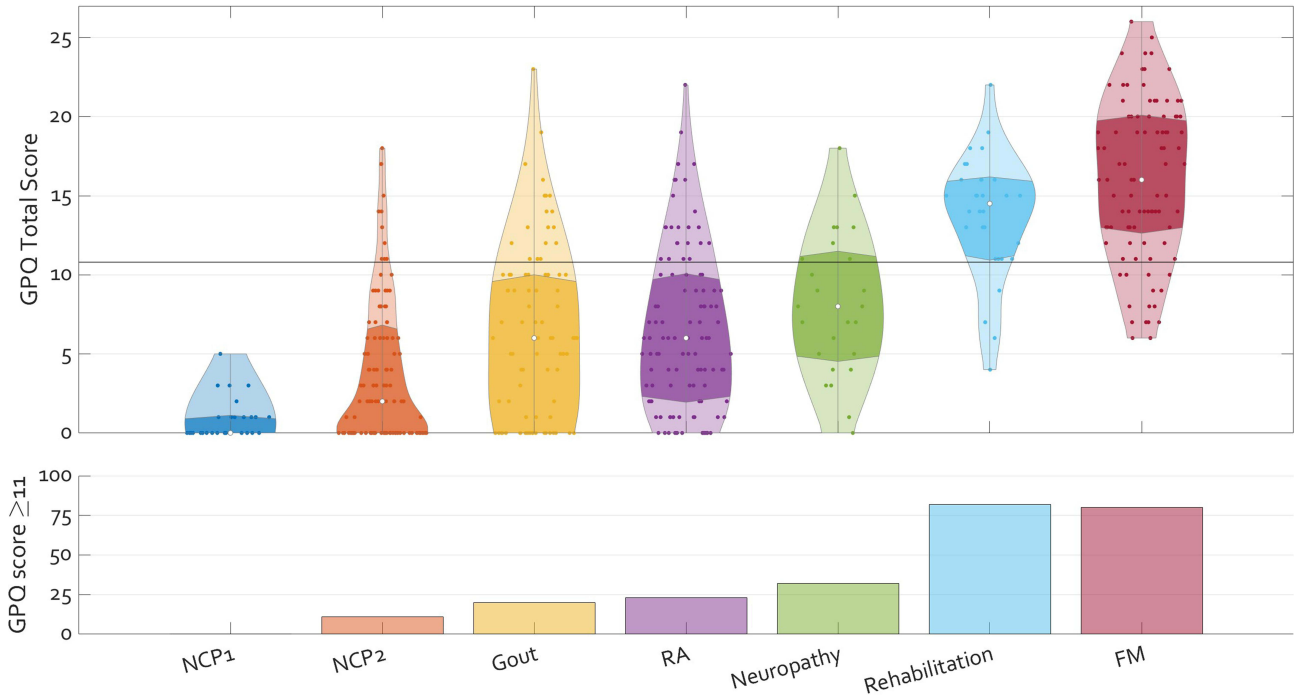

Figure 2 Visualization of the total GPQ scores per dataset. In the top panel, per dataset a violin plot is shown. Here, the white dot indicates the median value, whereby the darker area surrounding the median represents the interquartile range. Each individual dot represents the score of one subject in the dataset. In the bottom panel, the percentage of subjects per dataset is shown whereby the GPQ total score is higher than 10. This cut-off value is suggested by[5] to identify possible generalized pain hypersensitivity.
Abbreviations: NCP1, Nonclinical Population; RA, Rheumatoid Arthritis; RHB, Rehabilitation; FM, Fibromyalgia.

tool in a clinical setting, these findings could also be considered acceptable. The absolute reliability measures, more specifically the minimal detectable change, can help interpret changes within-person in outcomes in longitudinal settings.

The measurement properties of the GPQ total score as provided are, however, dependent on the measurement properties of the individual item scores. The GPQ, consisting of only 7 items, is therefore especially sensitive to one or more of the individual items displaying poor measurement properties. To identify potential areas of improvement for the questionnaire, the reliability of the individual items was also explored in this study. Here, we found that the ICCs of the seven individual items ranged from the lowest being 0.41 to the highest being 0.72; see also Table 2. Although ICCs for individual items are usually lower than for total scores, one highly significant systematic difference in scores was observed for question 3 ($p < 0.001$), with also one near-statistically significant ($p = 0.07$) systematic difference for question 2. Based on the content of these items, no clear explanation can be found for these (near-) significant systematic differences. Future research should help clarify whether these findings are accidental or robust. Overall, these findings on the individual items identify directions of future research which might help to improve the overall measurement properties of the GPQ.

Lastly, the reliability outcomes as presented in this study might have been affected by the heterogeneity of the patient population included in the first phase of this study. In this first phase of the study, approximately half of these included patients (N≈23) were suspected of having a possible fibromyalgia phenotype of RA (see also Jansen et al[8] for more information). In these patients, as compared to typical RA patients, central rather than peripheral pain mechanisms are thought to predominantly drive the pain.[22–25] It is unknown if and how the inclusion of this patient group has affected the reliability estimates.

Preliminary Reference Values

Preliminary reference values were provided for seven available datasets. Three of these datasets included patients with a clinical diagnosis of rheumatoid arthritis, fibromyalgia or gout. These datasets, which also contain larger numbers of patients ($N > 97$) as compared to most other datasets, could be useful as preliminary reference values in future studies. We specifically

note that these could be used as "preliminary" reference values as all of the included datasets in this study – including also these three larger datasets – do not contain a sufficiently high number of subjects nor are representative enough to serve as normative reference values. In the other two datasets (the 'Rehabilitation' dataset and the "Neuropathy" dataset), the patient groups have a more heterogenous background and the number of included patients is small. While such small and heterogenous datasets may not provide much value serving as robust reference value for future studies, they do provide relevant information related to the face validity of the GPQ. For additional comparison, two datasets (NCP1 and NCP2) have been included containing a non-clinical population comprising healthcare employees. From the descriptive statistics of both these datasets, it seems that the dataset obtained from the hospital (NCP2) reported higher overall scores with also a larger variation as compared to the nonclinical population scores obtained from the rehabilitation center (NCP1). One important difference between these two datasets which could help explain these differences, was that for the NCP1 dataset only volunteers without (a history of) chronic pain were included, while in the NCP2 dataset no such inclusion criteria were set. Consequently, in the NCP2 dataset, respondents which had underlying (persistent) pain could also be included. Our data in fact suggest that a substantial portion of the respondents in this dataset had underlying (persistent) pain, with 34 (31%) respondents reporting a pain level of 4 or higher on average over the past four weeks. As such, this could well explain the higher GPQ scores found in the NCP2 dataset as compared to the NCP1 dataset.

Comparing across all seven datasets can provide relevant insights into the face validity of the GPQ. The GPQ is developed with the aim of evaluating the presence and intensity of generalized pain hypersensitivity. Generalized pain hypersensitivity, ie widespread (whole-body) hypersensitivity to pain, is widely thought to be the result of maladaptive central pain regulatory mechanisms[26] – with multiple gain-control mechanisms which could be responsible (eg see Treede[27]) for the phenomenon. For the GPQ to have face validity, the group-level scores should be the highest in datasets wherein a high prevalence and intensity of generalized pain hypersensitivity could be expected to be present. Vice versa the GPQ scores should be lowest in the groups where no or little generalized pain hypersensitivity is expected to be present.

The lowest GPQ scores were indeed found in the non-clinical populations where no or only a very low prevalence of generalized pain hypersensitivity was expected. The highest GPQ scores were found in the fibromyalgia patients and the patients participating in a pain rehabilitation program. These findings correspond well with the expectations. For patients diagnosed with fibromyalgia, generalized pain hypersensitivity is considered one of the key characteristics of the disease.[28] For the rehabilitation patients, an earlier study found that 85% of the patients had multisite pain,[29] an important characteristic of generalized pain hypersensitivity. In this study, a very similar prevalence of 82% was found. In-between the nonclinical populations and the FM and RHB groups are the patient samples, which have been diagnosed with rheumatoid arthritis and gout, and the heterogenous neuropathy group. These findings are also as expected, given that in patients diagnosed with RA and gout, the pain in most patients originates from inflammatory (peripheral) processes rather than central processes, whereby in a portion of the patients with both RA and gout, also central pain regulatory mechanisms could be expected.[30–32] This corresponds well with the findings in this study, whereby 20% (Gout) and 23% (RA) has likely generalized pain. Overall, these findings provide strong arguments in favor of the GPQ having face validity as a measure of generalized pain hypersensitivity.

Limitations and Recommendations

For the evaluation of the test–retest reliability, a limitation of this study is that a subgroup of the RA patients underwent QST measurements prior to filling in the GPQ measurement for the first session. While we consider it unlikely that the QST measurements affected the GPQ results, such an effect cannot be ruled out. A limitation related to the reference values as presented here, is that the overall sample sizes are (very) small and that the included patients were not well characterized. Therefore, as a recommendation for future research, robust normative values should be obtained in larger, representative and well-characterized (ie not only in terms of diagnosis but also on eg psychological status, drug usage or treatments) patient groups.

Conclusion

This study shows that the GPQ has an acceptable reliability to be used as a tool to evaluate the presence and intensity of generalized pain hypersensitivity. The preliminary reference values reported here support the face validity of the questionnaire and can aid in the interpretation of future studies. From the analysis of individual items of the GPQ, directions for future

research are outlined which could further improve the measurement properties of the questionnaire. Taking the findings of this study together with other studies conducted with the GPQ, the GPQ seems promising for evaluating the presence and intensity of generalized pain hypersensitivity in chronic pain patients.

Data Sharing Statement

The dataset of the experiments reported here is available upon request to the corresponding author (Niels Jansen) of this paper.

Ethical Approval

All experiments were approved by the Medical research Ethics Committees United (MEC-U; NL73282.100.20) and in accordance with the 1964 Helsinki Declaration and its later amendments.

Code Availability

Code required to perform the analyses reported here is available upon request to the corresponding author of this paper.

Consent for Publication

All authors have approved of the manuscript and agreed to submit to Journal of Pain Research.

Consent to Participate

All participants provided written informed consent and were rewarded for participation in the experiment.

Funding

This work was supported by internal funding from the University of Twente and Medisch Spectrum Twente.

Disclosure

Prof. Dr. Harald Vonkeman reports grants from Galapagos, Boehringer Ingelheim; personal fees from Novartis, Pfizer, UCB, and Janssen, outside the submitted work. The authors declare that they have no other conflicts of interest in this work.

References

1. Loeser JD, Treede R-D. The Kyoto protocol of IASP basic pain terminology. *Pain*. 2008;137(3):473–477. doi:10.1016/j.pain.2008.04.025
2. Arendt-Nielsen L, Morlion B, Perrot S, et al. Assessment and manifestation of central sensitisation across different chronic pain conditions. *Eur J Pain*. 2018;22(2):216–241. doi:10.1002/ejp.1140
3. Nijs J, George SZ, Clauw DJ, et al. Central sensitisation in chronic pain conditions: latest discoveries and their potential for precision medicine. *The Lancet Rheumatology*. 2021;3(5):e383–e392. doi:10.1016/S2665-9913(21)00032-1
4. Ruscheweyh R, Marziniak M, Stumpenhorst F, Reinholz J, Knecht S. Pain sensitivity can be assessed by self-rating: development and validation of the Pain Sensitivity Questionnaire. *Pain*. 2009;146(1–2):65–74. doi:10.1016/j.pain.2009.06.020
5. van Bemmel PF, Oude Voshaar MA, ten Klooster PM, Vonkeman HE, van de Laar MA. Development and preliminary evaluation of a short self-report measure of generalized pain hypersensitivity. *J Pain Res*. 2019;12:395–404. doi:10.2147/JPR.S182287
6. Schmidt-Wilcke T, Clauw DJ. Fibromyalgia: from pathophysiology to therapy. *Nat Rev Rheumatol*. 2011;7(9):518–527. doi:10.1038/nrrheum.2011.98
7. Schmidt-Wilcke T, Diers M. New insights into the pathophysiology and treatment of fibromyalgia. *Biomedicines*. 2017;5(2):22. doi:10.3390/biomedicines5020022
8. Jansen N, ten Klooster PM, Vonkeman HE, van den Berg B, Buitenweg JR. Further evaluation of inflammatory and noninflammatory aspects of pain in rheumatoid arthritis patients. *Rheumatol Adv Pract*. 2023;7(3):rkad076. doi:10.1093/rap/rkad076
9. Polit DF. Getting serious about test–retest reliability: a critique of retest research and some recommendations. *Qual Life Res*. 2014;23(6):1713–1720. doi:10.1007/s11136-014-0632-9
10. Freynhagen R, Baron R, Gockel U, Tölle TR. Pain DETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr. Med. Res. Opin.* 2006;22(10):1911–1920. doi:10.1185/030079906X132488
11. Ten Klooster PM, Kraiss JT, Munters R, Vonkeman HE. Generalized pain hypersensitivity and associated factors in gout. *Rheumatology*. 2021;61(9):3640.
12. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. *Behav Res Ther*. 2006;44(1):1–25. doi:10.1016/j.brat.2005.06.006

13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
14. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1(1):30. doi:10.1037/1082-989X.1.1.30
15. Portney LG. *Foundations of Clinical Research: Applications to Evidence-Based Practice*. FA Davis; 2020.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
17. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J R Stat Soc Ser a Stat Soc*. 1983;32(3):307–317.
18. Harvill LM. Standard error of measurement: an NCME instructional module on. *Educ Meas Issues Pract*. 1991;10(2):33–41. doi:10.1111/j.1745-3992.1991.tb00195.x
19. Geerinck A, Alekna V, Beaudart C, et al. Standard error of measurement and smallest detectable change of the sarcopenia quality of life (sarqol) questionnaire: an analysis of subjects from 9 validation studies. *PLoS One*. 2019;14(4):e0216065. doi:10.1371/journal.pone.0216065
20. Tampin B, Bohne T, Callan M, et al. Reliability of the English version of the painDETECT questionnaire. *Curr Med Res Opin* 2017;33(4):741–748. doi:10.1080/03007995.2017.1278682
21. Kregel J, Vuijk PJ, Descheemaeker F, et al. The Dutch Central Sensitization Inventory (CSI): factor analysis, discriminative power, and test-retest reliability. *Clin J Pain*. 2016;32(7):624–630. doi:10.1097/AJP.0000000000000306
22. Lee YC, Lu B, Edwards RR, et al. The role of sleep problems in central pain processing in rheumatoid arthritis. *Arthritis Rheum*. 2013;65(1):59–68. doi:10.1002/art.37733
23. Lee YC. Effect and treatment of chronic pain in inflammatory arthritis. *Curr Rheumatol Rep*. 2013;15(1):300. doi:10.1007/s11926-012-0300-4
24. Boyden SD, Hossain IN, Wohlfahrt A, Lee YC. Non-inflammatory causes of pain in patients with rheumatoid arthritis. *Curr Rheumatol Rep*. 2016;18(6):30. doi:10.1007/s11926-016-0581-0
25. Lee YC, Nassikas NJ, Clauw DJ. The role of the central nervous system in the generation and maintenance of chronic pain in rheumatoid arthritis, osteoarthritis and fibromyalgia. *Arthritis Res Ther*. 2011;13(2):1–10. doi:10.1186/ar3306
26. Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. *Pain*. 2011;152(3 Suppl):S2–15. doi:10.1016/j.pain.2010.09.030
27. Treede R-D. Gain control mechanisms in the nociceptive system. *Pain*. 2016;157(6):1199–1204. doi:10.1097/j.pain.0000000000000499
28. Wolfe F, Smythe HA, Yunus MB, et al. The American college of rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheum*. 1990;33(2):160–172. doi:10.1002/art.1780330203
29. Köke A, Smeets R, Schreurs K, et al. Dutch dataset pain rehabilitation in daily practice: content, patient characteristics and reference data. *Eur. J. Pain*. 2017;21(3):434–444. doi:10.1002/ejp.937
30. Adami G, Gerratana E, Atzeni F, et al. Is central sensitization an important determinant of functional disability in patients with chronic inflammatory arthritides? *Ther Adv Musculoskelet*. 2021;13:1759720X2199325. doi:10.1177/1759720X21993252
31. Guler MA, Celik OF, Ayhan FF. The important role of central sensitization in chronic musculoskeletal pain seen in different rheumatic diseases. *Clin Rheumatol*. 2020;39(1):269–274. doi:10.1007/s10067-019-04749-1
32. Kieskamp SC, Paap D, Carbo MJG, et al. Central sensitization, illness perception and obesity should be considered when interpreting disease activity in axial spondyloarthritis. *Rheumatology*. 2021;60(10):4476–4485. doi:10.1093/rheumatology/keab019