RESEARCH

Assessing physical workload among people with musculoskeletal disorders: validity and reliability of the physical workload questionnaire

Lise Grethe Kjønø[1*], Rikke Munk Killingmo[1], Ørjan Nesse Vigdal[1], Margreth Grotle[1,2] and Kjersti Storheim[1,2]

Abstract

Background: Demands of physical work are related to musculoskeletal disorders, and hence, important to assess. The Physical Workload Questionnaire (PWQ) is based on 26 items related to physical workload. The PWQ has been translated into Norwegian, but its psychometric properties have not yet been tested. The aim of this study was to assess the validity and reliability of the PWQ among patients with musculoskeletal disorders.

Methods: A cross-sectional study with a test-retest design was conducted to assess construct validity (structural validity and hypothesis testing) and reliability (internal consistency and test-retest reliability) among employed patients with musculoskeletal disorders. Exploratory factor analysis was performed to assess the structural validity and number of items to be included in the Norwegian version of the PWQ. Hypothesis testing was assessed by 14 a priori hypotheses ("known" group, convergent and discriminant validity). Internal consistency was evaluated using Cronbach's alpha and test-retest reliability by Intraclass Correlation Coefficient ($ICC_{2.1}$), Standard Error of Measurement ($SEM_{agreement}$) and Smallest Detectable Change ($SDC_{95\%}$ind).

Results: In total, 115 patients with a mean age (SD) of 46 (9) years were included, of which 48 were included in the reliability analyses. Exploratory factor analysis resulted in two subscales: "Heavy physical work" (15 items, range 0–100) and "Long-lasting postures and repetitive movements" (7 items, range 0–100). No floor or ceiling effects were seen in the subscales. Twelve of the 14 (85%) predefined hypotheses were confirmed. The internal consistency with Cronbach's alpha was 0.94 and 0.85 on subscales 1 and 2, respectively. Test-retest reliability analyses demonstrated an $ICC_{2.1}$ of 0.96 (95% CI 0.88, 0.98) and 0.92 (95% CI 0.81, 0.96), SEM of 6.9 and 10.0 and $SDC_{95\%}$ind of 19.2 and 27.7 of subscales 1 and 2, respectively.

Conclusions: The Norwegian version of the PWQ demonstrated good validity and reliability and can be used to evaluate physical workload in patients with musculoskeletal disorders.

Keywords: Physical workload questionnaire, PWQ, Physical workload, Musculoskeletal disorders, Validity, Reliability

Background

The demands of physical work are associated with the development of various musculoskeletal disorders [1, 2], and to labour market participation [3, 4]. Work-related musculoskeletal disorders are the leading cause of sickness absence in Europe [5, 6]. According to existing

*Correspondence: lisegret@oslomet.no
[1] Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway
Full list of author information is available at the end of the article

literature, major risk factors for work-related musculoskeletal disorders are heavy lifting, working with a bent or twisted back or elevated arms, repetitive movements, and vibration [1, 2].

To prevent and reduce work-related musculoskeletal disorders, it is necessary to assess physical workload at the workplace [7]. The Physical Workload Questionnaire (PWQ) was developed by Bot et al. [8], with the aim of creating a short and simple self-report questionnaire for assessing physical workload in occupational health care and epidemiological research. Twenty-six items that were expected to have an association with either upper or lower extremity complaints were tested for dimensionality, internal consistency, and construct validity in a population with upper- and lower extremity musculoskeletal disorders in the Netherlands. The items formed two subscales and the results supported the internal structure, internal consistency, and construct validity [8], suggesting that PWQ is useful for assessing physical workload in a population with musculoskeletal disorders. To the best of our knowledge, the PWQ has not been translated into any other languages or been tested for its psychometric properties in other studies, hence there is a need for assessment of the questionnaire in a different population and among patients with a broader range of musculoskeletal disorders. The PWQ was previously translated and cross-culturally adapted into Norwegian according to international guidelines [9, 10]. The present study aims to test the PWQ's validity and reliability in terms of structural validity, hypothesis testing, internal consistency, and test-retest reliability among Norwegian patients with various musculoskeletal disorders.

Methods
This study was designed and performed in accordance with the COSMIN checklist [11] and guidelines for PROMs [12].

Design
We used a cross-sectional design, including a test-retest assessment.

Translation and cross-cultural adaptation
The translation and cross-cultural adaptation were done according to international guidelines [9, 10]. Two translators (one philologist and one clinician), whose mother tongue is Norwegian, independently translated the 26 items into Norwegian and synthesized them into one Norwegian version before it was translated back to English. Two translators and native English speakers, blinded the original PWQ items, independently performed the backtranslation and synthesized the two versions into one English version. An expert committee consisting of

the translators and two researchers from the research group (MG, RMK) reviewed the translations and agreed on a prefinal version. Ten patients with musculoskeletal disorders reviewed the prefinal Norwegian version. The items and responses were confirmed to be relevant and understandable without any proposed alterations. Since the prefinal version was acceptable and easy to comprehend, no changes were made for the final version.

Participants
Participants were recruited from an outpatient rehabilitation clinic in Akershus, Norway, between November 2015 and January 2018. Eligible participants were patients with different types of musculoskeletal disorders, aged 18 or above, working or on sick leave, who were referred to a specialist assessment and rehabilitation at the outpatient rehabilitation clinic. Exclusion criteria were patients being unable to speak, read or write in Norwegian. Inclusion was performed by clinicians, primarily physiotherapists, meeting with patients at the clinic. At baseline, all patients received written and oral information about the study, and provided their signed, informed consent.

According to recommended quality criteria by Terwee et al. [12] and Kline [13] we planned to recruit a minimum of 100 patients. These criteria suggest a minimum of 100 participants for assessing internal consistency, at least 50 participants for assessing reliability and floor or ceiling effects [12], and at least 4–10 participants for each item included in factor analysis [13].

Procedures and measurements
At baseline, patients completed the PWQ as part of a comprehensive questionnaire which also included sociodemographic variables, pain localization, intensity and history, psychosocial work environment, productivity costs and health-related quality of life.

The McGill pain drawing was used to measure pain localisation during the last week [14]. The Numeric Rating Scale (NRS) (range 0–10, a higher score indicates more severe pain) was used to measure average pain intensity in the last week [15]. The General Nordic Questionnaire for psychological and social factors at work (QPSnordic) was used to measure characteristics of the psychosocial work environment [16]. The iMTA Productivity Cost Questionnaire (iPCQ) was used to measure work status (occupation, paid job, working days/hours a week, sick leave and rehabilitation/work disability) and productivity costs [17]. The Short Form 36 Health Status Questionnaire (SF-36) (range 0–100, higher score indicates better health- related quality of life) was used to measure health-related quality of life. In addition, the Mechanical Exposure Index (MEI) (range 0–24, higher

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 3 of 10

score indicates higher physical workload) was used to measure physical workload [18].

Patients consenting to participate in the test-retest part of the study filled out the PWQ and a global question recording change in work status at a second meeting, preferably within 1 week. Patients reporting "unchanged" work status were considered stable and included in the test-retest reliability analysis.

The physical workload questionnaire

The PWQ is a self-report questionnaire for assessing physical workload [8]. The questionnaire consists of 26 items assessing force, dynamic and static load, repetitive load, (uncomfortable) postures, sitting, standing, and walking. In the only previous study, assessing dimensionality, internal consistency, and construct validity among patients with upper and lower extremity musculoskeletal disorders in the Netherlands, factor analysis revealed two subscales- twelve items related to the first subscale "Heavy physical work" and six items related to the second subscale "Long-lasting postures and repetitive movements" [8]. The remaining eight items were excluded due to low loading or to similar loading on both subscales. Each item is scored on a 4-point Likert scale with the response options: "seldom or never" (0), "sometimes" (1), "often" (2), and "(almost) always" (3). Scoring is conducted by adding up the responses to each item to produce a raw score. The final scores are calculated by dividing the raw score by the maximum possible score on the subscale, multiplied by 100, resulting in a final score ranging between 0 (no workload) and 100 (highest workload) for each subscale [8]. The Norwegian version of the 26 items on the PWQ is shown in Additional file 1.

Analysis

All data analyses were performed using SPSS version 26 (IBM Corporation, Armonk, NY, USA). The structural validity was explored using Exploratory Factor Analysis (EFA) based on the same 26 items which formed the basis of the study of Bot et al. [8]. The suitability of data for factor analysis was confirmed using the Kaiser-Meyer-Olkin measure of sampling adequacy (values above 0.6 considered acceptable), a significant Bartlett's Test of Sphericity and inspection of the correlation matrix (correlation coefficients of .3 and above preferable) [19]. Principal Component Analysis (PCA) was used to extract the factors followed by oblique rotation of factors using oblimin rotation. The number of factors to be retained was guided by three decision rules: Kaiser's criterion, retention of eigenvalues above 1, Cattel's scree plot [20], and by the use of Horn's parallel analysis [21]. To aid in the interpretation of the retained factors, we computed factor loadings after direct oblimin rotation, allowing factors

to correlate [19]. The next step involved interpreting the rotated solution by identifying which items loaded on each retained factor. Items with factor loading below 0.5 [22] and communalities value below 0.3 were excluded [23]. Items which cross-loaded were retained in the factor they loaded most strongly.

Hypothesis testing was assessed by 14 a priori hypotheses; "known" group validity (eight), convergent validity (two) and discriminant validity (four). The "known" group hypothesis are identical to those in the original study. They were tested with the same procedure as in the study of Bot et al., where it was hypothesised that physical workload would vary among different occupational groups [8]. As in the original study, the occupations of all included patients were classified into four groups based on expected physical load, and the subscale scores of the occupational groups were compared.

- Group 1: no physical load (for example teacher, manager)
- Group 2: heavy physical load (for example nurse, childcare worker)
- Group 3: long-lasting postures and repetitive movements (for example cashier, civil servant, engineer)
- Group 4: both heavy physical load and long-lasting postures and repetitive movements (for example electrician, farmer, mechanic)

Two investigators (LGK, ØNV) made the classifications independently, based on available occupation descriptions [24, 25]. Disagreements were resolved in a consensus meeting with a third investigator (RMK). Three occupations could not be classified (police, shop assistant and service employee) due to considerable physical workload variability within the occupations, and patients with these occupations were therefore excluded from the hypothesis analyses.

To assess convergent validity, both subscales were validated against the MEI [18]. The MEI includes similar questions to the PWQ, especially regarding heavy physical workload. We therefore expected high correlation between the MEI and the "Heavy physical work" subscale and moderate to high correlation between the MEI and the "Long-lasting postures and repetitive movements" subscale. To assess discriminant validity of the PWQ subscales, we formulated hypotheses regarding two dimensions from SF-36; "physical function" and "general health" [26]. These dimensions measure different constructs to the PWQ. We therefore expected low correlation between both PWQ subscales and the SF-36 dimensions. If >75% of the predefined hypotheses were confirmed, construct validity was considered acceptable [12]. Mann-Whitney U tests and Wilcoxon signed ranks

Kjønø et al. BMC Musculoskeletal Disorders   (2022) 23:282

Page 4 of 10

tests were used in "known" group analyses. Spearman's rho was used in all correlation analyses (convergent and discriminant validity) because the scales were not normally distributed. Correlation coefficients under 0.3, between 0.3 and 0.6 and over 0.6 were considered low, moderate and high, respectively [27]. The hypotheses are listed in Table 3.

The internal consistency of the subscales was examined using Cronbach's alpha. Cronbach's alpha between 0.70 and 0.95 gave a positive rating [12]. The item-total correlation was examined and items with values below 0.3 were excluded [28].

For test-retest assessment, a paired t-test was used to assess the mean difference between test and re-test. An intraclass correlation coefficient ($ICC_{2,1}$) was used to assess relative reliability. The acceptable level of ICC was set to ≥0.70 [12]. Absolute reliability (measurement error) was evaluated by standard error of measurement (SEM) and smallest detectable change (SDC). $ICC_{2,1}$ and

$SEM_{agreement}$ were used to account for the systematic difference between test and re-test [28]. SEM was estimated from the SPSS VARCOMP analysis; $SEM_{agreement} = \sqrt{(o^2_o + o^2_{po,e})}$, where o^2_o is the variance due to systematic error between observations and $o^2_{po,e}$ is the random error. Based on this, the SDC was estimated using the formula $SDC_{95\%}ind = 1.96 \times \sqrt{2} \times SEM_{agreement}$ [28].

Proportions of missing data and floor and/or ceiling effects were described. Floor or ceiling effects were considered to be present if more than 15% of patients reported either the lowest or the highest possible score [12].

Results

A total of 115 patients with a mean (SD) age of 46 (9) were included. Study sample characteristics are presented in Table 1. The majority of the patients were women. Almost all patients (90%) were in paid work, and more than half had been on sick leave during the

Table 1 Patient demographic characteristics and clinical status at baseline

	Cross sectional study sample		Test-retest study sample	
	$n =$		$n =$	
Female, N (%)	115	79 (68.7)	48	30 (62.5)
Age in years, mean (SD)	115	45.6 (9.3)	48	46.2 (8.6)
Mother tongue Norwegian, N (%)	115	100 (87)	48	43 (89.6)
Educational level high, N (%)	115	67 (58.2)	48	30 (62.5)
Work status, N (%)				
Employed or self-employed (paid job)	115	104 (90.4)	48	42 (87.5)
Working hours per week, median (range)	111	37.5 (7.5–52)	47	37.5 (7.5–52)
Working days per week, median (range)	102	5 (2–7)	41	5 (2–6)
Sick leave during last 4 weeks	114	79 (68.7)	48	33 (68.8)
Rehabilitation, work disability	115	7 (6.1)	48	4 (8.4)
Pain location, N (%)				
Lower extremities		70 (60.9)		27 (56.3)
Back		80 (69.6)		36 (75.0)
Neck		57 (49.6)		17 (35.4)
Upper extremities		55 (47.8)		15 (31.3)
Pain duration in months, median (range)	89	8.3 (0.2–360)	38	8.6 (0.2–360)
≤ 3 months, N (%)		13 (14.6)		5 (13.2)
> 3 months, N (%)		76 (85.4)		33 (86.8)
Pain severity in the last week (NRS 0–10), mean (SD)	112	5.2 (2.0)	48	4.9 (2.2)
Physical Workload Questionnaire (PWQ 0–100) Heavy physical work, median (range)	107	15.6 (0–82.2)	48	16.7 (2.2–82.2)
Long-lasting postures and repetitive movements, mean (SD)	110	49.1 (25.6)	48	49.2 (25.4)
Mechanical exposure index (MEI 0–24), median (range)	111	4 (0–20)	47	4 (0–20)
Health-related quality of life (SF-36 0–100), median (range)				
Physical function	115	75 (30–100)	48	70 (30–95)
General health	115	57 (5–97)	48	68.5 (15–97)

NRS Numeric rating scale (0 = no pain, 10 = worst possible pain), *PWQ* Physical Workload Questionnaire (0 = no workload, 100 = highest workload), *MEI* Mechanical Exposure Index (0 = no workload, 24 = highest workload), SF-36: The short form 36 Health status questionnaire (0 = maximum disability, 100 = no disability)

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 5 of 10

previous 4 weeks. On average, patients reported moderate pain severity, the majority had pain for more than 3 months, and the most frequently reported pain area was the back region. Physical workload was generally low.

Sixty-two patients completed the retest questionnaire, of which 48 reported no change in working conditions and had complete PWQ scores and could thus be included in the test-retest analysis. Patients participating in the test ($n = 115$) and the retest ($n = 48$) were largely similar, however, individuals included in the retest had slightly different pain site locations, physical function, and general health on the SF-36. The median (range) time interval between test and retest was 3 days.

Structural validity

Inspection of the correlation matrix revealed the presence of many coefficients above 0.3. Bartlett's Test of Sphericity was highly significant ($p < 0.001$), and the Kaiser-Meyer-Olkin measure of sampling adequacy value of 0.86 supported the factorability of the correlation matrix [19]. PCA revealed the presence of five factors with eigenvalues exceeding 1 (Kaiser's criterion), explaining 39, 16, 6, 4 and 4% of the variance, respectively. However, the results of Horn's parallel analysis indicated only two factors appropriate for retention and the scree plot suggested either a three- or two-factor solution, therefore, both the three- and two-factor solutions were inspected.

The three-factor solution explained a total of 61% (39, 16 and 6%) of the variance. Examination of the factor loadings revealed only two items in factor 3 ("sitting for long periods of time" and "visual display units (VDU) work for long periods of time", and, as a subscale in a questionnaire should be comprised of least three items [28], the three-factor solution was rejected. Therefore, PCA with oblimin rotation was repeated, forcing two factors. The items "sitting for long periods of time" (2) and "VDU work for long periods of time" (3) loaded highly negative on the first factor and below 0.5 on the second factor and were therefore excluded. Item 22 "climbing stairs" was excluded due to negative loading on factor 2 and loading below 0.5 on factor 1. Item 21 "operating peddles with your feet" showed low communalities value (0.276), indicating a poor fit with the other items in the factor, and was therefore excluded. Results from the three- and two-factor solutions are presented in Additional files 2 and 3.

Finally, a forced two-factor analysis with oblimin rotation on the remaining items was found to explain 58% (41 and 17%) of the total variance (Table 2). The items that loaded highly on the first factor were related to heavy physical work and the items that loaded highly on the second factor were related to static postures or repetitive movements. As a result, 22 items remained (15 items in factor 1 and 7 items in factor 2). The factor labels proposed by Bot et al. [8] suited the extracted factors in this analysis which resulted in: subscale 1 "Heavy physical work" and subscale 2 "Long-lasting postures and repetitive movements".

Construct validity

In total, 12 (85%) of the 14 predefined hypotheses were confirmed (Table 3), indicating acceptable construct validity.

Internal consistency

The Cronbach alpha value was 0.94 and 0.85 for subscale 1 and 2, respectively. The item-total correlation was 0.53–0.84 and 0.52–0.73 for subscale 1 and 2 respectively, indicating that all items correlated well with the total subscales.

Test-retest reliability

Relative and absolute reliability values for patients reporting no change in physical workload are presented in Table 4. Both subscales showed acceptable relative reliability ($ICC_{2,1} > 0.7$).

The proportion of missing data was very small, under 4% for all single items (ranging from 0.9 to 3.5%). For subscales 1 and 2, missing data was 7% and 4,3%, respectively. There were no floor or ceiling effects for any of the subscales. However, there were ceiling effects on five single items and floor effects on all but two single items. The highest floor effect was 79,8% (Additional file 4).

Discussion

In this study, the validity and reliability of the PWQ were found to be good when assessed in a sample with various musculoskeletal disorders.

Since the population in our study was made up of patients with various musculoskeletal disorders, and therefore different from the population in the original study (mainly upper- and lower extremity musculoskeletal disorders) [8], all 26 questions were included in the factor analyses. Factor analysis revealed that the PWQ could be divided into two subscales with a total of 22 items remaining in the final PWQ when tested in a sample with various musculoskeletal disorders. The present results were in line with those of the original study in terms of the number of subscales obtained and the nature of the items comprising each of the subscales. However, the number of items included in each subscale differed. In the current study, 3 additional items were included in subscale 1 ("working with vibrating tools", "walking on irregular surfaces" and "working in a twisted posture for long periods of time"), and one more item ("working in uncomfortable postures") was included in subscale 2.

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 6 of 10

Table 2 Final factor loadings after forced two factor solution and exclusion of items

Item	Pattern coefficients		Structure coefficients	
Does your work involve …	Factor 1	Factor 2	Factor 1	Factor 2
17. Physical hard work?	**0.84**		**0.86**	
13. Moving loads (more than 5 kg)?	**0.83**		**0.83**	
12. Work (ing) with your hands below knee level?	**0.80**		**0.79**	
11. Work (ing) with your hands above shoulder level?	**0.78**		**0.80**	
23. Squatting often?	**0.78**		**0.74**	
14. Moving heavy loads (more than 25 kg)?	**0.77**		**0.77**	
15. Exerting force with your arms or hands?	**0.75**	0.30	**0.80**	0.42
5. Kneeling or squatting for long periods of time?	**0.75**		**0.75**	
4. Walking long periods of time?	**0.73**		**0.70**	
25. Sitting or moving on your knees?	**0.71**		**0.74**	
16. Exerting maximal force?	**0.70**		**0.74**	
1. Standing for long periods of time?	**0.66**		**0.64**	
24. Walking on irregular surfaces?	**0.61**		**0.64**	
7. Working in a twisted posture for long periods of time?	**0.60**	0.46	**0.61**	**0.56**
20. Working with vibrating tools?	**0.57**		**0.58**	
6. Making the same movement for long periods of time for long periods of time?		**0.83**		**0.82**
18. Working in the same position for long periods of time?	−0.37	**0.82**		**0.76**
10. Holding your wrist in a bent or twisted position for long periods of time?		**0.75**		**0.77**
26. Doing repetitive tasks with arms, hands or fingers many times per minute?		**0.72**		**0.69**
8. Holding your neck in a bent forward or twisted position for long periods of time?		**0.63**		**0.66**
9. Bending or twisting your neck often?	0.33	**0.62**	0.43	**0.67**
19. Working in uncomfortable postures?	0.39	**0.56**	0.49	**0.62**
Eigenvalue	9.15	3.66		
Variance explained	41.6%	16.7%		
Total variance explained	58.3%			

Factor loadings < 0.3 are removed. Factor loadings > 0.5 are given in bold

Considering that the current study also included patients with back pain, and that the most frequently reported pain area was the back region, this difference was not entirely unexpected. Back pain is in several previous studies found to be associated with risk factors such as twisted posture [29–31], working with vibrating tools [32] and uncomfortable postures/ awkward postures [29, 30, 33]. This might explain why these items loaded strongly enough to be included in the subscales in the present study. As the PWQ originally was composed for patients with either upper- or lower extremity disorders some items may also not be applicable for those with back pain. For example, the items "neck bent forward", "turning/bending neck" and "repetitive tasks arms/hands" are items more often associated with neck- and upper limb pain [1, 2].

Regarding "known" group validity, we found that the median values were statistically significantly different between occupational groups for six of the eight hypotheses. In line with Bot et al. [8] we found that the PWQ clearly distinguished between the subscale scores

of the occupational group with "long-lasting postures and repetitive movements" as this group scored low on the first subscale and high on subscale two. In addition, all hypotheses regarding occupations classified as "both physical heavy load and long-lasting postures and repetitive movements" were confirmed. In the "known" group analyses, a significance level of 0.01 was chosen to adjust for multiple testing and give more power to the results thereby. However, when decreasing the level of significance, the probability of wrongly accepting the null hypothesis increases, thus increasing the possibility of type 2 error. The hypotheses regarding convergent validity and discriminant validity were confirmed. The MEI, which is a questionnaire assessing mechanical exposure of the shoulder-neck region [18], showed high correlation with the "heavy physical work" subscale and moderate correlation with the "long-lasting postures and repetitive movements" subscale. The SF-36 dimensions "physical function" and "general health" [26] measure constructs other than physical workload, and, as expected, low correlation was found with both subscales. Eighty-five

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 7 of 10

Table 3 Construct validity: A priori formulated hypotheses (*n* = 115)

	Value	Hypothesis confirmed
The mean score on the subscale "heavy physical work" is significantly higher for the occupational group with heavy physical load (*n* = 10) than for the occupational group with long-lasting postures and repetitive movements (*n* = 52)	*p* = 0.008	Yes
The mean score on the subscale "heavy physical work" is significantly higher for the occupational group with heavy physical load (*n* = 10) than for the occupational group without physical load (*n* = 20)	*p* = 0.038	No
The mean score on the subscale "long-lasting postures and repetitive movements" is significantly higher for the occupational group with long-lasting postures and repetitive movements (*n* = 53) than for the occupational group with heavy physical load (*n* = 10)	*p* = 0.001	Yes
The mean score on the subscale "long-lasting postures and repetitive movements" is significantly higher for the occupational group with long-lasting postures and repetitive movements (*n* = 53) than for the occupational group without physical load (*n* = 22)	*p* < 0.001	Yes
The mean score on the subscale "heavy physical work" is significantly higher for the occupational group with both heavy physical load and long-lasting postures and repetitive movements (*n* = 14) than for the occupational group without physical load (*n* = 20)	*p* < 0.001	Yes
The mean score on the subscale "long-lasting postures and repetitive movements" is significantly higher for the occupational group with both heavy physical load and long-lasting postures and repetitive movements (*n* = 13) than for the occupational group without physical load (*n* = 22)	*p* = 0.001	Yes
In the occupational group with heavy physical load (*n* = 10), the mean score on the subscale "heavy physical work" is significantly higher than on the subscale "long-lasting postures and repetitive movements"	*p* = 0.173	No
In the occupational group with long-lasting postures and repetitive movements (*n* = 52), the mean score on the subscale "long-lasting postures and repetitive movements" is significantly higher than on the subscale "heavy physical work"	*p* < 0.001	Yes
High correlation between "heavy physical work" score and the MEI score (*n* = 104)	rho = 0.783 *p* < 0.001	Yes
Moderate to high correlation between "long-lasting postures and repetitive movements" score and the MEI score (*n* = 106)	rho = 0.402 *p* < 0.001	Yes
Low correlation between "heavy physical work" score and "physical function" score (SF-36) (*n* = 107)	rho = -0.288 *p* = 0.003	Yes
Low correlation between "long-lasting postures and repetitive movements" score and "physical function" score (SF-36) (*n* = 110)	rho = -0.155 *p* = 0 .107	Yes
Low correlation between "heavy physical work" score and "general health" score (SF-36) (*n* = 107)	rho = -0.126 *p* = 0.195	Yes
Low correlation between "long-lasting postures and repetitive movements" score and "general health" score (SF-36) (*n* = 110)	rho = -0.282 *p* = 0.003	Yes

Hypotheses were analysed with Mann-Whitney U tests, Wilcoxon signed ranks tests (both presented with *p* values) and correlation analysis (presented with Spearman's rho and *p* values)

MEI Mechanical Exposure Index, *SF-36* Short Form 36 Health Status Questionnaire

Table 4 Test-retest reliability

	First test	Re-test	Difference	ICC$_{2,1}$ (95% CI)	SEM$_{agreement}$	SDC$_{95\%}$ ind
Heavy physical work *(n = 48)*	27.3 (24.7)	22.2 (21.9)	−5.1 (8.5)*	0.96 (0.88, 0.98)	6.9	19.2
Long-lasting postures and repetitive movements *(n = 48)*	49.2 (25.4)	42.7 (24.1)	−6.5 (12.6)*	0.92 (0.81, 0.96)	10.0	27.7

ICC Intraclass correlation coefficient, *SEM* Standard error of measurement, *SDC* Smallest detectable change

Score values and difference values are presented with mean (SD)

*P values < 0.01

percent of the predefined hypotheses were confirmed, indicating acceptable construct validity [12].

The good internal consistency of the subscales indicated that the items in the respective subscales correlated well with each other, and thus that they measured the same concept [12]. However, a Cronbach's alpha value exceeding 0.9 may indicate that some items are redundant [22]. Examination of the item-total statistics showed that three items would decrease the Cronbach's alpha to 0.93 if they were removed from the scale

Kjønø *et al. BMC Musculoskeletal Disorders* (2022) 23:282

Page 8 of 10

("moving loads more than 5kg", "exerting force with your arms and hands" and "physical hard work"). However, the decrease was minimal, and we considered the items to be important and to contribute to the content validity of the instrument. This result is consistent with the original study [8], which showed a Cronbach's alpha of 0.92–0.93 on subscale 1 and 0.86–0.87 on subscale 2.

The ICC was well above the minimum standard of both subscales and was therefore considered to be acceptable, which suggests that the PWQ is a reliable measure in our population [28]. There was a statistically significant decrease in difference score from test to re-test in both subscales. However, the decrease may be considered to be low as the scale ranged from 0 to 100. The absolute reliability, presented as measurement error and reported in the actual scale unit, is more clinically useful than the decrease in difference score and relative reliability. The $SDC_{95\%ind}$ results indicate that a score of self-reported physical workload at the individual level would have to change by 19.2 and 27.7 on subscales 1 and 2, respectively, to ensure that the change was not a result of measurement error [28]. On a scale from 0 to 100, these values may indicate relatively large measurement error. Further research is needed to evaluate the responsiveness and MIC of the PWQ subscales.

Limitations

The main limitation of this study is aspects of the sample size. Although the sample size of 115 participants was above the minimum threshold for conducting factor analysis, it was rather low in regard to the number of subjects per variable according to the rules of thumb (4 to 10 subjects per variable) [13]. New guidelines from COSMIN, published after the data collection for this study was finished, require a minimum of 7 participants per item to be considered "very good" in the quality criteria [34]. Hence, the sample size of this study may have influenced the robustness of the factor analysis. Guidelines recommend a minimum of 50 participants in test-retest analyses [11, 12]. Although 62 patients participated in the test-retest study, only 48 could be included in the analyses and there might be some imprecision in our estimates regarding test-retest assessment. In addition, our sample were recruited from a clinic located in a wealthy city close to the capital of Norway, which may imply that this study sample consist of patients with high socioeconomic status. Previous studies showed that low socioeconomic status is associated with higher exposure of physical workload [35, 36], and there is reason to believe that by recruiting participants from a wider geographical area we would have reached a broader population regarding occupational variation. This might influence the degree of representativeness to other populations of people in work with musculoskeletal disorders, in particular those with low socioeconomic status and those who are seeking primary health care. A second potential limitation is that we included patients who were on sick leave, in which could potentially introduce recall bias to exposure estimates. Furthermore, the time interval between measurements may be another potential limitation. Test-retest reliability should be assessed in a stable population with an appropriate time interval between measurements [12]. In the current study the time interval was median 3 days (range 1–10), meaning it was shorter than recommended for many of the patients. There is a potential risk of recall bias if the interval between the test and the re-test is too short. However, we believe that the comprehensive questionnaire with a high number of questions used in the first test most likely reduced recall bias when the same questionnaire was filled out only a few days later. In addition, self-ratings may suffer from misclassification. There are studies showing that workers with musculoskeletal disorders may overestimate the physical load compared to healthy workers [37, 38]. Even when participants are motivated to report the workload accurately, they may have difficulties with recalling and accurately reporting the information [39], or that pain level at the day of answering the questionnaire affects the self-reported level of physical workload [40], which may threaten the validity of the questionnaire. We also have a lack of data on eligible patients who declined to participate.

A strength of this study is that we followed the COSMIN checklist and PROM guidelines in the assessment [11, 12]. The number of items was reduced in a systematic manner by performing EFA according to guidelines. Our study also added convergent and discriminant validity to the construct validity and is the first to assess test-retest reliability of the PWQ.

Conclusions

The PWQ, consisting of two subscales: "Heavy physical work" and "Long-lasting postures and repetitive movements", showed good validity and reliability when used among patients with long-lasting musculoskeletal disorders receiving rehabilitation in an outpatient clinic in Norway. This study indicates that the PWQ can be used in clinical and occupational healthcare and for research purposes among patients with musculoskeletal disorders. Further research should be conducted on hypothesis testing and test-retest reliability in other populations and clinical settings. As well, the clinical value of the PWQ in relation to work-related musculoskeletal disorders should be investigated.

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 9 of 10

Abbreviations

CI: Confidence Interval; COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments; EFA: Exploratory Factor Analysis; ICC: Intraclass Correlation Coefficient; iPCQ: The iMTA Productivity Cost Questionnaire; MEI: Mechanical Exposure Index; MIC: Minimal Important Change; NRS: Numeric Rating Scale; PCA: Principal Component Analysis; PROMs: Patient-Related Outcome-Measures; PWQ: Physical Workload Questionnaire; QPSnordic: General Nordic Questionnaire for Psychological and social factors at work; SD: Standard Deviation; SDC: Smallest Detectable Change; SEM: Standard Error of Measurement; SF-36: Short Form 36 Health Status Questionnaire; VDU: Visual Display Unit.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-022-05222-y.

Additional file 1. The Norwegian version of the PWQ.

Additional file 2. Table of Pattern and Structure matrix of three-factor solution after EFA with oblimin rotation.

Additional file 3. Table of Pattern and Structure matrix of two-factor solution after EFA with oblimin rotation.

Additional file 4 Table of missing data, floor- and ceiling effects of the PWQ subscales and items ($n = 115$).

Authors' contributions
LGK, MG and RMK designed the study. MG and KS contributed to the funding of the study. RMK took part in data collection and preparation of the datasets. LGK analysed the data. LGK, RMK, ØNV, MG and KS contributed to interpretation of the data. LGK drafted the manuscript, with all authors providing critical revision and approving the final manuscript.

Availability of data and materials
The datasets generated and analysed during the current study are not publicly available due to protection of the participants' anonymity but are available from mgrotle@oslomet.no upon reasonable request.

Declarations

Ethics approval and consent to participate
The study is considered a quality assessment project by the Norwegian Regional Committee for Medical and Health Research Ethics (reference no. 2014/1634/REK vest) and was approved by the Norwegian Centre for Research Data (reference no. 42149) in 2018. This study was performed in accordance with relevant guidelines and regulations and all participants provided written informed consent to participate in the study.

Consent for publication
Not applicable.

Competing interests
The authors declares that they have no competing interests.

Author details
[1] Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway. [2] Research and Communication Unit for Musculoskeletal Health (FORMI), Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.

References

1. Da Costa BR, Vieira ER. Risk factors for work-related musculoskeletal disorders: a systematic review of recent longitudinal studies. Am J Industrial Med. 2010;53:285–323. https://doi.org/10.1002/ajim.20750.
2. Mayer J, Kraus T, Ochsmann E. Longitudinal evidence for the association between work-related physical exposures and neck and/or shoulder complaints: a systematic review. Int Arch Occup Environ Health. 2012;85(6):587–603. https://doi.org/10.1007/s00420-011-0701-0.
3. Andersen LL, Fallentin N, Thorsen SV, Holtermann A. Physical workload and risk of long-term sickness absence in the general working population and among blue-collar workers: prospective cohort study with register follow-up. Occup Environ Med. 2016;73(4):246–53. https://doi.org/10.1136/oemed-2015-103314.
4. Sterud T. Work-related mechanical risk factors for long-term sick leave: a prospective study of the general working population in Norway. Eur J Public Health. 2014;24(1):111–6. https://doi.org/10.1093/eurpub/ckt072.
5. Bang CK, Lund T, Labriola M, Villadsen E, Bültmann U. The fraction of long-term sickness absence attributable to work environmental factors: prospective results from the Danish work environment cohort study. Occup Environ Med. 2007;64:487–9. https://doi.org/10.1136/oem.2006.028563.
6. Griffith LE, Shannon HS, Wells RP, Walter SD, Cole DC, Côté P, et al. Individual participant data meta-analysis of mechanical workplace risk factors and low back pain. Am J Public Health. 2012;102(2):309–18. https://doi.org/10.2105/AJPH.2011.300343.
7. Stock SR, Fernandes R, Delisle A, Vezina N. Reproducibility and validity of workers' self-reports of physical work demands. Scand J Work Environ Health. 2005;31(6):409–37. https://doi.org/10.5271/sjweh.947.
8. Bot SDM, Terwee C, van Der Windt DAWM, Feleus A, Bierma-Zeinstra SM, Knol DL, et al. Internal consistency and validity of a new physical workload questionnaire. Occup Environ Med. 2004;61:980–6. https://doi.org/10.1136/oem.2003.011213.
9. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993;46(12):1417–32. https://doi.org/10.1016/0895-4356(93)90142-N.
10. Beaton ED, Bombardier EC, Guillemin EF, Ferraz EMB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186–91. https://doi.org/10.1097/00007632-200012150-00014.
11. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49. https://doi.org/10.1007/s11136-010-9606-8.
12. Terwee CB, Bot SDM, de Boer MR, van Der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.
13. Kline P. The handbook of psychological testing. London: Routledge; 1993.
14. Öhlund C, Eek C, Palmblad S, Areskoug B, Nachemson A. Quantified pain drawing in subacute low Back pain: validation in a nonselected outpatient industrial sample. Spine. 1996;21(9):1021–30. https://doi.org/10.1097/00007632-199605010-00005.
15. Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. Spine. 2000;25(24):3140–51. https://doi.org/10.1097/00007632-200012150-00009.
16. Dallner M. Validation of the general Nordic questionnaire (QPSNordic) for psychological and social factors at work. Nordisk Ministerråd: København; 2000.

Kjønø *et al. BMC Musculoskeletal Disorders*   (2022) 23:282

Page 10 of 10

17. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Roijen LV-H. The iMTA productivity cost questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value Health. 2015;18(6):753–8. https://doi.org/10.1016/j.jval.2015.05.009.

18. Balogh I, Ørbaek P, Winkel J, Nordander C, Ohlsson K, Ektor-Andersen J. Questionnaire-based mechanical exposure indices for large population-reliability, internal consistency and predictive validity. Scand J Work Environ Health. 2001;27(1):41–8. https://doi.org/10.5271/sjweh.585.

19. Tabachnick BG, Fidell LS. Using multivariate statistics. 6th ed. Harlow: Pearson; 2014.

20. Cattell RB. The scree test for the number of factors. Multivar Behav Res. 1966;1(2):245–76. https://doi.org/10.1207/s15327906mbr0102_10.

21. Watkins M. Monte Carlo PCA for parallel analysis (computer software). State College: Ed & Psych Associates; 2000.

22. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill; 1994.

23. Pallant J. SPSS survival manual: a step by step guide to data analysis using IBM SPSS. 6th ed. Maidenhead: McGraw Hill; 2016.

24. National Center for O*NET Development. O*NET OnLine. http://www.onetonline.org/. Cited 15 Jan 2021.

25. Direktoratet for høyere utdanning og kompetanse. utdanning.no. https://www.utdanning.no. Cited 15 Jan 2021.

26. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473–83.

27. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81:15–20. https://doi.org/10.1053/apmr.2000.20619.

28. de Vet H, Terwee C, Mokkink L, Knol D. Measurement in medicine. Cambridge: Cambridge University Press; 2011.

29. Videman T, Ojajarvi A, Riihimaki H, Troup JD. Low back pain among nurses: a follow-up beginning at entry to the nursing school. Spine. 2005;30(20):2334–41. https://doi.org/10.1097/01.brs.0000182107.14355.ca.

30. Van Nieuwenhuyse A, Somville PR, Crombez G, Burdorf A, Verbeke G, Johannik K, et al. The role of physical workload and pain related fear in the development of low back pain in young workers: evidence from the BelCoBack study; results after one year of follow up. Occup Environ Med. 2006;63(1):45–52. https://doi.org/10.1136/oem.2004.015693.

31. Campo M, Weiser S, Koenig KL, Nordin M. Work-related musculoskeletal disorders in physical therapists: a prospective cohort study with 1-year follow-up. Phys Ther. 2008;88(5):608–19. https://doi.org/10.2522/ptj.20070127.

32. Bovenzi M, Schust M, Mauro M. An overview of low back pain and occupational exposures to whole-body vibration and mechanical shocks. La Medicina del lavoro. 2017;108:419–33. https://doi.org/10.23749/mdl.v108i6.6639.

33. Bovenzi M. Metrics of whole-body vibration and exposure-response relationship for low back pain in professional drivers: a prospective cohort study. Int Arch Occup Environ Health. 2009;82(7):893–917. https://doi.org/10.1007/s00420-008-0376-3.

34. Mokkink L, Prinsen CB, Patrick D, Alonso J, Bouter L, Terwee C. COSMIN risk of Bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171–9. https://doi.org/10.1007/s11136-017-1765-4.

35. Mehlum IS, Kristensen P, Kjuus H, Wergeland E. Are occupational factors important determinants of socioeconomic inequalities in musculoskeletal pain? Scandinavian journal of work. Environ Health. 2008;34(4):250–9. https://doi.org/10.5271/sjweh.1269.

36. Aittomäki A, Lahelma E, Roos E, Leino-Arjas P, Martikainen P. Gender differences in the Association of age with physical workload and functioning. Occup Environ Med. 2005;62(2):95–100. https://doi.org/10.1136/oem.2004.014035.

37. Balogh I, Ørbæk P, Ohlsson K, Nordander C, Unge J, Winkel J, et al. Self-assessed and directly measured occupational physical activities—influence of musculoskeletal complaints, age and gender. Appl Ergon. 2004;35(1):49–56. https://doi.org/10.1016/j.apergo.2003.06.001.

38. Hansson G-Å, Balogh I, Byström J, Ohlsson K, Nordander C, Asterland P, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands.

Scand J Work Environ Health. 2001;(1):30–40. https://doi.org/10.5271/sjweh.584.

39. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211–7. https://doi.org/10.2147/JMDH.S104807.

40. Viikarijuntura E, Rauas S, Martikainen R, Kuosma E, Riihimaki H, Takala E-P, et al. Validity of self-reported physical work load in epidemiologic studies on musculoskeletal disorders. Scand J Work Environ Health. 1996;22(4):251–9. https://doi.org/10.5271/sjweh.139.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.