

Validating Self-Report Measures of Pain and Function in Patients Undergoing Hip or Knee Arthroplasty

Paul W. Stratford, Moneet Dogra, Linda Woodhouse, Deborah M. Kennedy, Greg F. Spadoni

ABSTRACT

Purpose: To investigate the factorial and construct validity of a four-item pain intensity scale, the P4, in patients awaiting primary total hip or knee arthroplasty secondary to osteoarthritis.

Method: A construct validation design was applied to a sample of convenience of 117 patients (mean age 65.6 [SD = 11.2] years) at their preoperative visit. All patients completed the P4 and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Exploratory and confirmatory factor analyses were used to examine the factorial structure of the P4 and WOMAC. To evaluate construct validity, we examined the correlation between the P4 and WOMAC pain sub-scales and the ability of the P4 to differentiate between patients awaiting hip and knee replacement.

Results: Two distinct factors consistent with the themes of pain and function were identified with P4 and WOMAC physical function items, but not with the WOMAC pain and physical function items. The P4 correlates more with the WOMAC pain scores ($r = 0.67$) than with the WOMAC physical function scores ($r = 0.60$).

Conclusion: The P4's validity was supported in this patient group. The use of the P4 with the WOMAC physical function sub-scale provides a more distinct assessment of pain and function than the WOMAC pain and physical function scales.

Keywords: measurement, osteoarthritis, pain, physical function, total joint arthroplasty, validity

Stratford PW, Dogra M, Woodhouse L, Kennedy DM, Spadoni GF. Validating self-report measures of pain and function in patients undergoing hip or knee arthroplasty. Physiother Can. 2009;61;189-194.

RÉSUMÉ

Objectifs : Étudier la validité factorielle et conceptuelle d'une échelle d'intensité de la douleur comportant quatre paliers, la P4, chez les patients en attente d'une arthroplastie primaire de la hanche ou du genou résultant d'une arthrose.

Méthode : Nous avons eu recours à un modèle de validité conceptuelle pour un échantillon de commodité composé de 117 patients (moyenne [SD]; âge 65,6 [11,2] ans) en consultation préopératoire. Tous les patients ont complété le P4 et l'indice d'arthrose de WOMAC (universités Western Ontario et McMaster). Des analyses factorielles exploratoires et confirmatives ont été utilisées pour étudier la structure factorielle de l'échelle P4 et de l'indice de WOMAC. Dans le but d'évaluer la validité conceptuelle, la corrélation entre la sous-échelle de douleur P4 et l'indice de WOMAC et la capacité de P4 de différencier les patients en attente d'un remplacement de la hanche de ceux en attente d'un remplacement du genou ont été examinées.

Résultats : Deux facteurs distincts cohérents avec les thèmes de la douleur et de la fonction ont été identifiés grâce aux éléments de l'échelle P4 et de l'indice de WOMAC touchant la fonction, mais non avec l'élément douleur et fonction de l'indice de WOMAC. L'échelle P4 est davantage en corrélation avec les scores portant sur la douleur de l'indice de WOMAC ($r = 0,67$) qu'avec les scores portant sur la fonction physique du même indice ($r = 0,60$).

Conclusions : La validité de l'échelle P4 a été appuyée dans ce groupe de patients. L'utilisation conjointe de l'échelle P4 et de la sous-échelle de fonction physique de l'indice de WOMAC permet donc une évaluation plus distincte de la douleur et de la fonction qu'avec les échelles de douleur et de fonction de l'indice de WOMAC.

Mots clés : arthroplastie complète de l'articulation, arthrose, douleur, fonction physique, validité des mesures

Paul W. Stratford, PT, MSc: Professor, School of Rehabilitation Science, and Associate Member, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario; Scientific Affiliate, Department of Surgery, Sunnybrook Holland Orthopaedic & Arthritic Institute, Sunnybrook Health Sciences Centre, Toronto, Ontario.

Moneet Dogra, PT, MSc: Advanced Practice Physiotherapist, Sunnybrook Holland Orthopaedic & Arthritic Centre, Toronto, Ontario.

Linda Woodhouse, PT, PhD: Assistant Professor, School of Rehabilitation Science, Faculty of Health Sciences, McMaster University, Hamilton, Ontario; Scientific Affiliate, Departments of Orthopaedics and Rehabilitation, Hamilton Health Sciences, Hamilton, Ontario; Scientific Affiliate, Department of Surgery, Sunnybrook Holland Orthopaedic & Arthritic Institute, Sunnybrook Health Sciences Centre, Toronto, Ontario.

Deborah M. Kennedy, PT, MSc: Manager of Hip and Knee Program Development, Sunnybrook Holland Orthopaedic & Arthritic Centre, Toronto Ontario; part-time Assistant Clinical Professor, School of Rehabilitation Science, McMaster University, Hamilton, Ontario.

Gregory F. Spadoni, PT, MSc, FCAMT: Assistant Professor, School of Rehabilitation Science, McMaster University, Hamilton, Ontario; Director, Peak Performance Physiotherapy, Hamilton, Ontario.

Address correspondence to *Paul W. Stratford*, School of Rehabilitation Science, Institute of Applied Health Sciences (4th Floor, Rm 403), McMaster University, 1400 Main St. West, Hamilton, ON L8S 1C7 Canada. Tel.: 905-525-9140 ext. 22523. Fax: 905-524-0069. E-mail: stratfor@mcmaster.ca.

DOI:10.3138/physio.61.4.189

INTRODUCTION

Pain and limitations of physical function are two concerns of patients with osteoarthritis (OA) of the hip or knee seeking care from physical therapists. In turn, decreasing pain and increasing lower-extremity functional status are two important treatment goals. Although pain is currently assessed by self-report only, functional status can be evaluated by both self-report and performance measures. Consistent with the view that pain and physical function are important outcomes, the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group identified pain, physical function, the patient's global rating of change, and imaging examination as essential core outcomes for patients with OA.¹ Interestingly, this group suggested that self-reports of physical function are obligatory, while performance-based tests are optional.¹

One of the most frequently cited self-report measures used to evaluate patients with OA of the hip or knee is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).² The WOMAC defines lower-extremity physical function as "the ability to move around."³ A concern identified in recent years is that patients' responses to self-report measures intended to provide unique information on lower-extremity functional status are influenced not only by their ability to move around but also by the pain they experience when moving around.⁴ Although it is likely that this phenomenon has an impact on most, if not all, self-report measures of physical function, this unwanted association appears to be magnified by the WOMAC.^{4,5} A likely explanation for the inextricable link between WOMAC pain and physical function scores is the similarity of item content and phrasing on the pain and physical function sub-scales.⁶ The impact of this association is problematic, in that WOMAC physical function scores have been shown to reflect little or no change in patients whose time to complete performance activities (e.g., 40 m walk, stair test, timed up-and-go test) more than doubled.⁴

In retrospect, the strong and partially spurious association between WOMAC pain and physical function scores is not surprising, given the consistency with which factor analyses of this measure have uniformly failed to identify factors consistent with the themes of pain and function.⁷⁻⁹ The results of these studies have shown that factor themes are more consistent with activities (e.g., a factor would contain the items *pain with walking* and *difficulty with walking*) than with the independent themes of pain and function.

Given the results of these factor analyses, we aimed to determine whether a non-activity-centred pain scale, namely the P4,¹⁰ coupled with the WOMAC physical function items could achieve factorial validity. Our primary goal was to assess whether P4 and WOMAC

physical function items yielded factors consistent with the themes of pain and function, retrospectively, when applied to patients with OA of the hip or knee awaiting total joint arthroplasty. A secondary purpose was to estimate the extent to which P4 scores demonstrated cross-sectional construct validity on this patient sample by examining (1) the correlation between P4 and WOMAC pain scores (convergent validity); (2) the ability of P4 scores to correlate more highly with WOMAC pain scores than with WOMAC physical function scores (discriminant validity); and (3) the ability of P4 scores to identify higher pain levels in patients awaiting total hip arthroplasty (THA) than in patients awaiting total knee arthroplasty (TKA) (i.e., known group validity).

METHODS

Patients

Participants in the current study represent the initial sample (all patients at the time of data analysis) from an ongoing observational study being conducted at a tertiary-care orthopaedic facility in Toronto. Designated a Centre of Excellence for hip and knee replacement, the Sunnybrook Holland Orthopaedic & Arthritic Institute is one of the largest-volume arthroplasty sites in the country.

Patients were eligible to participate in this study if they met the following inclusion criteria: diagnosis of OA, scheduled for primary TKA or THA; sufficient language skills to communicate in written and spoken English; and absence of neurological, cardiac, and psychiatric disorders and of other medical conditions that would significantly compromise physical function. Our sample size was one of convenience rather than being based on a formal sample-size calculation. Ethics approval for the study was received from the institution's research ethics review board, and all participants provided written informed consent.

Design

We applied a cross-sectional construct validation design. Patients completed the P4 and full WOMAC (pain, stiffness, and physical function sub-scales) at a single time point prior to THA or TKA.

Measures

P4

The P4 is a four-item instrument measuring pain intensity.¹⁰ The items inquire about pain in the morning, in the afternoon, in the evening, and with activity. Each item is scored on an 11-point numeric pain scale (0 = no pain, 10 = pain as bad as it can be). Item scores are summed to yield a total score from 0 to 40.

WOMAC

Patients completed all components (pain, stiffness, physical function) of the LK 3.1 version of the WOMAC;³ the present study used data from the pain and physical function sub-scales only. The WOMAC pain sub-scale consists of 5 items, and the WOMAC physical function sub-scale consists of 17 items. All WOMAC items are scored on a scale from 0 to 4; higher scores represent more pain or difficulty. Maximum possible scores for the pain and physical function sub-scales are 20 and 68, respectively.

Analysis

We applied confirmatory factor analysis¹¹ to assess whether the four items composing the P4 formed a single factor. We conceptualized a one-factor measurement model with uncorrelated error terms. The model fit was assessed by applying the Comparative Fit Index (CFI), Relative Fit Index (RFI), Tucker-Lewis Index (TLI), root mean-square error of approximation (RMSEA), and the model fit chi-square and associated *p*-value.¹¹ Although no single standard exists to define an acceptable fit for a model, the following values are generally accepted: CFI, RFI, and TLI values exceeding 0.95 indicate good fit; RMSEA values below 0.05 represent good fit, and values below 0.08 indicate reasonable fit; and a *p*-value > 0.05 associated with the model fit chi-square.^{11,12}

We calculated Cronbach's alpha to assess the internal consistency of both pain measures and computed the standard error of measurement (SEM_{IC}) to evaluate the precision associated with each measure's score. The SEM_{IC} was calculated as follows:

$$SD\sqrt{1 - \alpha}$$

where *SD* = standard deviation of the P4.

A principal components analysis¹³ was applied with Promax rotation to explore the factorial structure when P4 pain items and WOMAC physical function items were examined in the same analysis. For purposes of comparison, we performed a similar analysis using the WOMAC pain and physical function items.

We estimated the convergent construct validity of the P4 by correlating its scores with those of the WOMAC pain sub-scale. The discriminant validity of the P4 was estimated by contrasting Pearson's correlation coefficient between the P4 and the WOMAC pain sub-scale with the correlation between the P4 and the WOMAC physical function sub-scale. We calculated two-sided 95% confidence intervals (CI) for all correlation estimates and for the difference in pain scores for patients awaiting THA and TKA. Analyses were performed using SPSS version 15 (SPSS Inc., Chicago, IL).

RESULTS

Complete P4 and WOMAC data were available for 117 of the 120 eligible patients presenting for assessment; two of these 117 patients did not respond to all WOMAC physical function items. Of the 117 patients, 63 were awaiting TKA and 66 were female. Of the 66 female patients, 41 were awaiting TKA. The sample's mean (SD) age and body mass index were 65.6 (11.2) years and 30.4 (6.3) kg/m², respectively. Table 1 summarizes pain and function scores by joint site. Two patients had the lowest possible P4 pain score of 0, and three patients had the highest possible pain score of 40.

The confirmatory factor analysis results supported the premise that P4 items assess a single concept. CFI, RFI, and TLI all exceeded 0.98; the RMSEA was 0.07, and the model fit chi-square was 3.23 (*p* = 0.20). The factor loadings were as follows: morning, 0.84; afternoon, 0.95; evening, 0.95; and activity, 0.75.

The internal consistency of the P4 was 0.93; the SEM_{IC} was 2.7 P4 points. Results of the principal components analysis for the P4 and WOMAC physical function items are displayed in Table 2. As illustrated, the P4 items group as a separate factor consistent with the theme of pain. Results of the principal components analysis for the WOMAC pain and physical function sub-scales are found in Table 3. This analysis clearly demonstrates that the WOMAC pain items do not form a distinct factor but, rather, are aligned with their respective physical function activities.

Table 4 presents correlations among P4, WOMAC pain, and WOMAC physical function scores. The convergent validity of the P4 is supported to the extent that for the correlation between the P4 and WOMAC pain scores, *r* > 0.50. The discriminant validity of the P4 is supported to the extent that P4 scores are more highly correlated with WOMAC pain scores than with WOMAC physical function scores.

Comparison of pain intensity between patients awaiting THA and TKA is shown in Table 1. Although the point estimate of pain was 2.1 P4 pain points greater (95% CI: -1.7–5.9) for patients awaiting total hip

Table 1 Pain and Physical Function Summary Scores by Site of Problem

Measure	Mean (SD)
P4	
Knee (<i>n</i> = 63)	20.6 (10.8)
Hip (<i>n</i> = 54)	22.7 (9.6)
Combined (<i>n</i> = 117)	21.6 (10.3)
WOMAC Pain	
Knee (<i>n</i> = 63)	8.3 (4.0)
Hip (<i>n</i> = 54)	8.8 (4.0)
Combined (<i>n</i> = 117)	8.5 (4.0)
WOMAC Physical Function	
Knee (<i>n</i> = 62)	26.0 (12.7)
Hip (<i>n</i> = 53)	32.8 (13.2)
Combined (<i>n</i> = 115)	29.1 (13.3)

Table 2 WOMAC and P4 Physical Function Pattern Matrix Factor Loadings

Items	Factors		
	1	2	3
WOMAC			
Difficulty socks on	0.99	-0.28	0.01
Difficulty socks off	0.99	-0.17	-0.05
Difficulty in/out bed	0.78	-0.02	0.08
Difficulty bending to floor	0.75	-0.01	0.09
Difficulty on/off toilet	0.64	0.28	-0.06
Difficulty in/out car	0.59	0.28	-0.02
Difficulty lying in bed	0.59	0.04	0.14
Difficulty light duties	0.55	0.28	0.04
Difficulty heavy duties	0.41	0.37	-0.23
Difficulty up from sitting	0.36	0.31	0.12
Difficulty down stairs	-0.38	0.99	-0.01
Difficulty up stairs	-0.09	0.78	0.13
Difficulty shopping	0.28	0.55	0.00
Difficulty standing	0.25	0.54	0.05
Difficulty sitting	0.38	0.51	-0.09
Difficulty walking	0.32	0.48	0.05
Difficulty in/out bathtub	0.28	0.46	0.00
P4			
Pain evening	0.01	-0.04	0.96
Pain afternoon	0.07	-0.09	0.96
Pain morning	0.16	-0.03	0.84
Pain activities	-0.21	0.31	0.78

Table 3 WOMAC Pain and Physical Function Pattern Matrix Factor Loadings

Items*	Factors			
	1	2	3	4
Difficulty socks off	0.96	0.11	-0.04	-0.22
Difficulty socks on	0.94	0.17	-0.00	-0.35
Difficulty bending to floor	0.80	0.09	-0.10	0.05
Difficulty in/out bed	0.64	-0.13	0.32	0.05
Difficulty heavy duties	0.52	-0.01	-0.20	0.34
Difficulty in/out car	0.50	0.03	0.19	0.20
Difficulty on/off toilet	0.49	-0.12	0.35	0.21
Difficulty light duties	0.49	0.17	0.14	0.11
<i>Pain walking</i>	-0.08	0.82	0.13	0.00
<i>Pain standing</i>	0.07	0.80	0.06	0.01
Difficulty standing	0.29	0.76	-0.27	0.12
Difficulty walking	0.24	0.66	0.03	-0.02
Difficulty shopping	0.16	0.50	0.16	0.09
<i>Pain night in bed</i>	-0.14	0.10	0.99	-0.16
Difficulty lying in bed	0.21	-0.14	0.80	-0.02
<i>Pain sit or lying</i>	-0.06	0.28	0.64	0.02
Difficulty sitting	0.14	0.25	0.41	0.13
Difficulty down stairs	-0.24	0.06	-0.04	0.92
Difficulty up stairs	0.01	0.19	-0.07	0.72
Difficulty in/out bathtub	0.43	-0.07	-0.17	0.61
<i>Pain stairs</i>	-0.32	0.47	0.11	0.59
Difficulty up from sitting	0.34	-0.23	0.18	0.53

*Pain sub-scale items in italics; physical function sub-scale items in roman type.

Table 4 Pearson's Correlation Coefficients among P4, WOMAC Pain, and WOMAC Physical Function Scores

	P4 <i>r</i> (95% CI)	WOMAC Pain <i>r</i> (95% CI)
WOMAC Pain <i>r</i> (95% CI)	0.67 (0.56–0.76)	–
WOMAC Physical Function <i>r</i> (95% CI)	0.60 (0.47–0.70)	0.76 (0.67–0.83)

arthroplasty, this difference was not statistically significant, as the confidence interval of the difference included zero.

DISCUSSION

Our goal was to examine the construct validity of the P4 when applied to patients with OA of the hip or knee awaiting total joint replacement. We were particularly interested in determining whether factorial validity could be established when P4 items were combined with WOMAC physical function items. Our results showed that pain and function items loaded on separate factors. To determine whether this finding was an artefact of our data set, we also examined the factorial structure of WOMAC pain and function items. Consistent with the findings of previous factor analyses,^{7–9} our results failed to support a factorial structure of the WOMAC that discriminates pain from physical function items. This study suggests that a non-activity-focused pain scale coupled with an activity-focused physical function scale can achieve factorial validity.

In addition to factorial validity, the present study also examined three other aspects of construct validity. We examined convergent construct validity by correlating P4 scores with WOMAC pain scores, obtaining a correlation of 0.67. It has been our experience that measures intended to assess the same concept often correlate above 0.70; previous investigations of the P4 have reported correlations of approximately 0.85 with the numeric pain rating scale.¹⁰ There are several possible explanations for our lower-than-expected correlation. The first is chance: given that the upper confidence limit on our point estimate of 0.67 is 0.76, it may well be that this study's estimate is lower than the true population value. A second explanation is that the time frame associated with the reporting of pain differs between the P4 and the WOMAC: the P4 inquires about pain over the past two days, whereas the WOMAC asks about pain over the past week. The third explanation focuses on the nature of the items: the WOMAC asks about pain with specific activities, whereas the P4 inquires about pain at specific times of day (morning, afternoon, evening) and with activity. Finally, given the WOMAC's inability to demonstrate factorial validity, it is likely that WOMAC pain scores are influenced by patients' functional status as well as by the pain they are experiencing.

Our study also supported the discriminant validity of the P4 to the extent that its point estimate correlation with WOMAC pain scores was greater than its correlation with WOMAC physical function scores. By contrast, WOMAC pain scores showed a substantially greater correlation with WOMAC physical function scores than with P4 scores. Given the lack of factorial validity of WOMAC pain and function items, the high correlation between WOMAC pain and physical function scores is not

surprising. Discriminant validity is important because it provides evidence that a measure assesses what it is intended to assess rather than a general concept.

Our third validation construct was the premise that the P4 would display higher pain scores for patients awaiting THA than for those awaiting TKA. This premise was based on the findings of a previous study of patients with similar demographic and geographic characteristics to participants in the present study.¹⁴ Specifically, the previous study found that patients awaiting THA reported a mean WOMAC pain score of 9.1 (SD=3.4), compared to 7.8 (SD=2.7) for patients awaiting TKA.¹⁴ Although in the current study the point estimates of pain intensity for both the P4 and the WOMAC pain sub-scale indicated greater amounts of pain for patients awaiting THA, the differences included the value of zero (i.e., they were not statistically significant). One explanation for this finding is insufficient power. We calculated the power for the P4 contrast as 0.29 based on the following assumptions: a between-group difference (hip-knee) of 2.1, a pooled SD of 10.2, and a one-tailed test of significance at $p=0.05$. A second explanation is that our premise was incorrect. There is inconsistency in the literature as to the relative magnitude of preoperative pain scores for patients awaiting THA and TKA: in some reports, patients awaiting THA have higher pain scores, whereas in other studies those awaiting TKA have higher pain scores.^{14–16}

A secondary purpose of our study was to provide more information about the P4's measurement properties in the context of patients with OA of the hip or knee. Because measurements occur in context, their properties are specific to the measure's scores, not to the test or measure itself.¹⁷ Previous validation studies of the P4 included patients with a variety of orthopaedic problems attending outpatient physiotherapy clinics. In our study, the context was patients with OA of the hip or knee awaiting total joint arthroplasty. The Cronbach's alpha value of 0.93 reported in this study and the SEM_{IC} of 2.7 P4 points are consistent with estimates reported previously ($\alpha=0.92$, SEM_{IC}=2.8 P4 points).¹⁰ The confirmatory factor analysis fit indices reported here are also consistent with those reported previously.¹⁰

There are several limitations to our work. First, our study was based on a sample size of convenience rather than on an a priori sample-size calculation. Although our sample size of 117 patients provided reasonably narrow confidence intervals for many of the correlation coefficients, it is clear that this sample size would be underpowered for formal hypothesis testing. For example, applying the estimates of effect (2.1 P4 points) and variation (SD=10.2 P4 points) obtained from our study, a sample size of 293 patients per group would be required for a one-tailed Type I error of 0.05 and a power of 0.80. A second limitation of our study is that, as per the questionnaire design, WOMAC physical function items were

completed after completion of WOMAC pain items, meaning that responses to WOMAC physical function items were not free of influence imposed by responses to WOMAC pain items. Having acknowledged this, it is buoying to note that factorial validity was achieved for P4 and WOMAC physical function items. To remedy any response bias owing to WOMAC pain items, a direction for future inquiry would be to administer only the P4 and the WOMAC physical function sub-scale.

We view this study as the first step in a series of investigations of the P4 with patients undergoing THA or TKA. Clearly, establishing the extent to which a measure is valid and clinically useful is an ongoing process.¹⁷ Subsequent investigations will address several complementary topics, including (1) a cross-validation study of the factorial validity of the P4 and self-report function items; (2) cross-sectional and longitudinal convergent construct validity studies of P4 scores and change scores with those of other non-activity-specific pain measures, such as the numeric pain rating scale; and (3) estimation of the P4's responsiveness and of clinically important within-patient change in P4 points.

CONCLUSION

Our results show that factorial validity is better achieved when self-reports of pain and physical function are assessed using the P4 and the WOMAC physical function sub-scale together. Future investigations should address the limitations of this study related to sample size and to the temporal relationship of WOMAC pain and physical function responses.

KEY MESSAGES

What Is Already Known on This Subject

It is widely accepted that we need reliable and valid clinical tools to evaluate outcomes in patients with osteoarthritis (OA). One such measure, the WOMAC, was developed to evaluate self-reported pain, physical function, and stiffness in patients with lower-extremity OA. Despite its widespread use, studies have repeatedly questioned the factorial validity of the WOMAC in evaluating pain and physical function. Using the WOMAC to evaluate both pain and physical function together presents challenges in terms of separating the constructs of pain and physical function in patients with lower-extremity OA.

What This Study Adds

This study demonstrates that the P4, a simple four-item questionnaire with a low respondent and therapist burden, measures a single concept: pain intensity. Further, it reaffirms that the WOMAC does not clearly distinguish between the constructs of self-reported pain

and self-reported physical function. Our findings suggest that it would be better to evaluate self-reported pain and physical function using a combination of the P4 questionnaire and the physical function sub-scale of the WOMAC, rather than the WOMAC alone. Furthermore, numerous studies have demonstrated that physical function must be evaluated using both self-report and physical performance measures. We posit that the same concept must be applied when evaluating pain. Future research should focus on developing and validating measures that quantify neurophysiologic pain in patients with lower-extremity OA awaiting total joint arthroplasty.

REFERENCES

- Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. *J Rheumatol.* 1997;24:799–802.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–40.
- Bellamy N. WOMAC Osteoarthritis Index user guide IV. Brisbane: University of Queensland; 2000.
- Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol.* 2006;59:160–7.
- Parent E, Moffet H. Comparative responsiveness of locomotor tests and questionnaires used to follow early recovery after total knee arthroplasty. *Arch Phys Med Rehabil.* 2002;83:70–80.
- Stratford PW, Kennedy DM. Does parallel item content on WOMAC's pain and function subscales limits its ability to detect change in functional status? *BMC Musculoskelet Disord* [serial on the Internet]. 2004 Jun 9 [cited 2009 Aug 11];5(17) [9 p.]. Available from: <http://www.biomedcentral.com/1471-2474/5/17>.
- Kennedy D, Stratford PW, Pagura SMC, Wessel J, Gollish JD, Woodhouse LJ. Exploring the factorial validity and clinical interpretability of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). *Physiother Can.* 2003; 55:160–8.
- Thumboo J, Chew LH, Soh CH. Validation of the Western Ontario and McMaster University Osteoarthritis Index in Asians with osteoarthritis in Singapore. *Osteoarthr Cartilage.* 2001;9:440–6.
- Faucher M, Poiradeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test–retest reliability and construct validity of the WOMAC and Lequesne indexes. *Osteoarthr Cartilage.* 2002;10: 602–10.
- Spadoni GF, Stratford PW, Solomon PE, Wishart LR. Development and cross-validation of the P4: a self-report pain intensity measure. *Physiother Can.* 2003;55:32–8.
- Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum; 2001.
- Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. Mahwah, NJ: Lawrence Erlbaum; 1996.
- Norman GR, Streiner DL. Biostatistics: the bare essentials. 2nd edn. Hamilton, ON: BC Decker; 2000.
- Stratford PW, Kennedy DM, Riddle DL. New study design evaluated the validity of measures to assess change after hip or knee arthroplasty. *J Clin Epidemiol.* 2009. Forthcoming.
- Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, *et al.* A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthr Cartilage* 2001;9:137–46.
- Fortin PR, Clarke AE, Joseph L, Liang MH, Tanzer M, Ferland D, *et al.* Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. *Arthritis Rheum.* 1999;42:1722–8.
- Messick S. Validity. In: Lin RL, editor. Educational measurement. 3rd edn. Phoenix, AZ: Oryx Press. 1993. p. 13.