Measuring pain intensity in patients with neck pain

Does it matter how you do it?

Kamper, Steven J.; Grootjans, Sanneke J.M.; Michaleff, Zoe A.; Maher, Christopher G.; Mcauley, James H.; Sterling, Michele

Published in:
Pain Practice

DOI:
10.1111/papr.12169

Licence:
Other

Title
Measuring pain intensity in patients with neck pain: Does it matter how you do it?

Authors
Steven J Kamper[1, 2]
Sanneke J M Grootjans[1]
Zoe A Michaleff[1]
Christopher G Maher[1]
James H McAuley[3]
Michele Sterling[4]

Affiliations
[1] The George Institute for Global Health, University of Sydney. Australia
[2] The EMGO+ Institute, VU University Medical Centre, Amsterdam. The Netherlands
[3] Neuroscience Research Australia
[4] Centre for National Research on Disability and Rehabilitation Medicine (CONROD), University of Queensland. Australia

Corresponding author
For all correspondence and reprints:
Steven Kamper
Address: The George Institute for Global Health, University of Sydney
PO Box M201 Missenden Rd
Camperdown NSW 2050
Phone: (61) 2 8238 2413
E-mail: skamper@george.org.au

No. Figures / Tables
1 / 5

Key words: Pain score, measurement, neck pain, whiplash.

Abstract

The aim of this study was to investigate whether variations in the way that pain intensity is measured in patients with neck pain influences the magnitude of pain ratings. The study uses data from three longitudinal studies (n= 361 at baseline) on people with neck pain due to whiplash injuries. Pain measures included verbal rating scales, numerical rating scales and a visual analogue scale. Different measures asked patient to rate; current pain, average pain over 24 hours, over 1 week, or over 4 weeks. Scores were converted to a 0-100 scale and tracked over time, correlations between measures were calculated. Mixed models regression was used to explore the factors which influenced the differences between scores on the measures. Scores on the different measures were significantly different from each other in each dataset ($p<0.02$). The effect of recall period was significant in all datasets and the effect of number of response options was significant in 2 of 3 datasets. Pain intensity ratings appear to be sensitive to method of measurement. It is likely the length of recall time (e.g. pain today versus average pain over 4 weeks) has a significant influence on pain ratings. The influence of number of response options is less certain. Systematic reviewers should not uncritically rescale and pool absolute pain scores from instruments with varying scale descriptors or recall periods.

Introduction

Measuring pain intensity

There are many ways researchers and clinicians measure pain intensity. Although the Visual Analogue Scale (VAS) and Numerical Rating Scale (NRS) are used most commonly in clinical research[1, 2], various verbal rating scales[3] (VRS) are also used often as part of larger, multi-dimensional outcome measures[4, 5]. These pain rating scales are known as 'subjective-' or 'patient reported outcomes', because they measure perception of pain as experienced by the patient. Such measures typically form the primary outcome in studies of painful conditions such as back pain[6].

There is a considerable heterogeneity in the way pain intensity outcomes are collected and reported in clinical research[7]. While studies have compared pain scores on different measures[8-10], further investigation into the features of those measures that influence pain ratings is necessary. Such features include the words used in the question, descriptors on the scale, the number of response options and the time period over which patients are asked to recall their pain. This can create issues of interpretability and comparability for readers of primary studies and for researchers conducting systematic reviews. Of particular relevance to meta-analyses is the question of whether it is appropriate for researchers to rescale any pain measures available in the primary studies to a common metric e.g. 0-100 scale for the purposes of pooling pain outcomes[11]. While many studies report correlations between scores on different scales, fewer seek to empirically investigate the nature of, and reasons for, systematic differences between scores on the scales[8]. To our knowledge this has never been done in patients with whiplash associated disorders (WAD).

Whiplash Associated Disorder is a common musculoskeletal condition which typically begins after a rear-end motor vehicle accident where acceleration-deceleration energy is transferred

to the neck[12]. The most common symptom of WAD, is neck pain (90-100% of patients). Pain intensity is the most consistently identified prognostic factor for poor outcome[11, 13] and clinical practice guidelines stress the need to measure patient's pain[14]. With the multitude of options available for scoring pain intensity, it is important for clinicians to be aware of differences in a patient's pain rating depending on the way the question is asked and which scale is used. If there are systematic differences in the way patients rate their pain, it is important for clinicians to be consistent with the selection of a measurement tool for patients with WAD. This could be important on an individual level when assessing a patient's progress and also for comparing patients with each other regarding their treatments and outcome, especially across multiple practices. Further, clinicians need to be aware of these differences in order to correctly interpret research findings and incorporate these findings into their own practice.

Aims

This study aims to investigate the following questions : (1) are some ratings of pain scored systematically higher than others, and (2) do the time period over which patients are asked to recall their pain or the number of response options systematically influence pain ratings.

Methods

Participants

This study involves secondary analysis of data collected as part of three clinical studies conducted in Sydney and Brisbane, Australia (Table 1). Study 1[15] was a longitudinal cohort study investigating the prognosis of acute WAD, Studies 2[16] and 3[17] were RCTs testing the effectiveness of exercise interventions in people with chronic WAD. The following

inclusion criteria were common to all studies; neck pain due to a car accident, age between 18 and 65 and fluency in written and spoken English. Participants were excluded if cervical scans showed fracture or dislocation or if they had a diagnosis of serious spinal pathology or major psychiatric illness.

The principle point of difference between the cohorts was with regard to the duration of symptoms on entry to the study. Participants in Study 1 were enrolled within 1 month of their car accident and were recruited from hospital emergency rooms, via newspaper advertisements and through referral from physiotherapy practices. Participants whose symptoms had persisted for greater than 3 months and less than 12 months (Study 2) and greater than 3 months but less than 5 years (Study 3) made up the chronic cohorts. Participants in Studies 2 and 3 were recruited via newspaper advertisement and from the records of the third party insurance administrator (Motor Accidents Authority, NSW and The Motor Accident Insurance Commission, Queensland).

Table 1 Sample and study characteristics

	Study 1	Study 2	Study 3
Age years (SD)	42.0 (13.4)	43.3 (14.7)	43.7 (12.9)
Gender % female	69.2	66.4	64.5
Mean symptom duration Days (SD)	19 (9)	285 (117)	456 (688)
Neck Disability Index % (SD)	36.4 (17.3)	38.0 (13.2)	36.2 (15.9)
Follow-Up points and Sample sizes (n)	Baseline (101)	Baseline (134)	Baseline (146)
	3 months (91)	6 weeks (132)	3 months (128)
	6 months (86)	-	6 months (121)
	12 months (89)	12 months (125)	12 months (104)

Measures

Assessments were carried out at baseline and at either two or three follow-up points in each study (Table 1). Data were collected in an assessment booklet containing various questionnaires and scales to assess socio–demographic variables (e.g. age, gender), pain severity, psychological measures, and disability. Verbal rating scales contained a written descriptor for each scale point (Likert scale), NRSs contained write descriptors at the end points and numbers for all other scale points, the VAS was a horizontal line with written descriptors at the end points. The same questionnaires were not available from each of the studies, the pain intensity questions available and used in the analyses are outlined in Table 2.

Table 2 Description of the pain measures

Measure	Question	Scale Type	Study 1	Study 2	Study 3
Neck Disability Index (NDI)	What is your pain intensity right now?	6-point VRS	Y	Y	Y
MOS Short Form 36 (SF36)	How much bodily pain did you have during the past 4 weeks?	6-point VRS	Y	Y	Y
Visual Analogue Scale (VAS)	What is your average pain intensity in the past 24 hours?	10cm VAS	Y	N	N
Numeric Rating Scale 24 hour (NRS24)	What is your average pain intensity in the last 24 hours?	11-point NRS	N	Y	Y
Numeric Rating Scale 1 week (NRSWk)	What is your average pain intensity in the last week?	11-point NRS	N	N	Y
Functional Rating Index (FRI)	What is your pain intensity right now?	5-point VRS	N	Y	N
Whiplash Disability Questionnaire (WDQ)	How much pain do you have today?	11-point NRS	N	N	Y

The Neck Disability Index (NDI) is a 10 item pain intensity and daily activity questionnaire that measures daily limitations after cervical spine injury[4]. Item 1 was extracted for this

study, asking the patient to rate 'pain intensity right now' on a 6-point verbal rating scale; 'no pain (0) to worst pain imaginable (5)'.

The Visual Analogue Scale (VAS) scale is a 10cm horizontal line, with extremes marked 'no pain' (left) and the 'worst pain imaginable' (right)[18]. Patients were asked to mark the spot on the line that best represents their pain intensity over the last 24 hours.

SF 36 is a health-related quality of life questionnaire which comprises 36 questions divided into eight domains[1, 5]. For this study, question 7 was used, asking 'how much bodily pain did you have during the past four weeks', it is rated on a 6-point verbal rating scale ranging from 'none (1) to very severe (6)'.

The Functional Rating Index (FRI) contains 10 items to measure disability associated with back and/or neck pain. For this study the item 'pain intensity right now' was extracted on a 5-point verbal rating scale, ranging from '0 = no pain' to '5 worst possible pain'.

The Numerical Rating Scale (NRS) scale has the same terminal anchors as a VAS but consists of numbers from 0-10[3]. Patients were asked to circle the number which best represents their pain. Two NRSs were used, one which asked about pain over the last 24 hours and the other about pain over the past week.

The Whiplash Disability Questionnaire (WDQ) is a modified version of the NDI with 13 items designed to evaluate WAD[19]. For this study the item 'how much pain do you have today' was extracted. Pain is scored on an 11-point NRS scale from 0 = no pain to 10 = worst pain imaginable.

All measures were coded for recall period (current pain, pain over 24 hours, pain over 1 week, pain over 4 weeks), for response option (5/6-point, 11-point, 10cm VAS) and for follow-up point (baseline, 3 months, 6 months, 12 months).

Data Analysis

In each cohort, at each follow up time point scores for pain measures were converted to a 0-100 scale by simple multiplication and plotted along with their 95% confidence intervals. VRS scores were converted by assigning the 'pain-free' rating a score of 0 and the highest rating a score of 100, other ratings were distributed evenly between these two endpoints. Bivariate correlations between scores on measures were calculated and Pearson's r reported. To compare the multiple means with each other, one-way analysis of variance (ANOVA) was performed at each time-point in each study (total of 11 ANOVAs). To explore the factors potentially responsible for systematic differences between measures linear mixed models regression was performed, with subject as a random factor, and recall period, response option and follow-up point as fixed factors. Recall period was coded as 1, immediate; 2, 24 hours; 3, one week; and 4; 4 weeks. Number of response options were coded as 1, 6-point, 2, 11-point, and 3, 10cm. Data were analysed using the SPSS 20.0 statistical program.

Results

Participants

The participants in the three studies were comparable in terms of age, gender balance and level of disability as assessed by total NDI score.

Are some pain measures scored systematically higher than others?

Scores on the three measures in the acute cohort (Study 1) were significantly correlated at all time points, Pearson's r for the correlations ranged from 0.23 to 0.80. Mean pain scores (Table 3) showed a consistent relationship over time, with the highest scores coming from the SF-36 Bodily Pain question, second NDI pain intensity item, and third the VAS scores. While mean pain levels fell over the study period and differences between the measures became

smaller, the pattern remained stable (Figure 1a). Results from the ANOVAs indicated that there were significant differences between scores on the different measures at all time points ($F = 4.9\text{-}26.3$, $df = 2$, $p = < 0.05$).

Scores on the three measures from the chronic cohort in Study 2 were significantly correlated at all time points, Pearson's r for the correlations ranged from 0.36 to 0.78. Mean pain scores (Table 3) showed large differences between some measures but not others. The highest ratings again came from the SF-36 and the lowest from the NDI, scores from the FRI item and NRS fell between these two but did not appear to be different from one another (Figure 1b). Results from the ANOVAs indicated that there were significant differences between scores on the different measures at all time points ($F = 13.1\text{-}19.3$, $df = 3$, $p = < 0.05$).

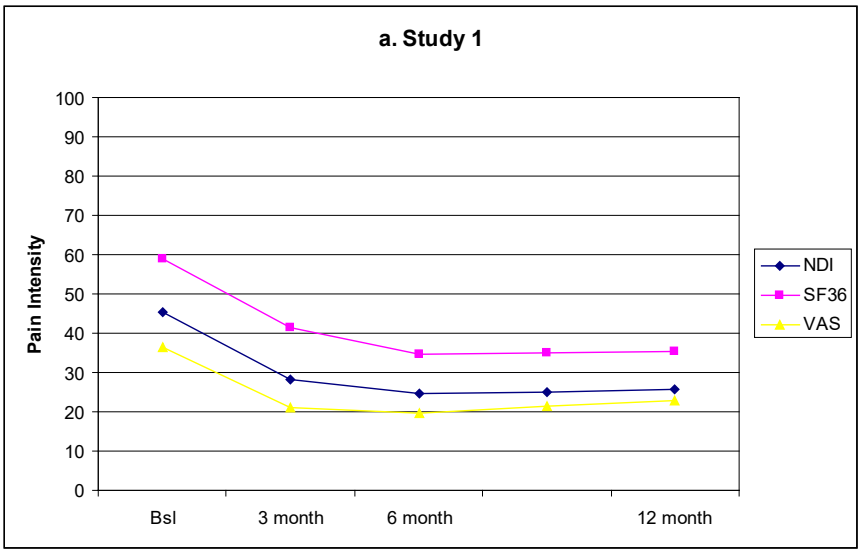
Scores on the three measures from the chronic cohort in Study 3 were significantly correlated at all time points, Pearson's r for the correlations ranged from 0.26 to 0.92 for all time points. Mean pain scores (Table 3) showed a similar pattern in that SF-36 and NDI item scores were again the highest and lowest respectively. However, there was a smaller difference between the means from all the measures in this cohort (Figure 1c). Results from the ANOVAs indicated that there were significant differences between scores on the different measures at all time points ($F = 4.1\text{-}12.1$, $df = 4$, $p = < 0.05$).

Factors influencing pain score

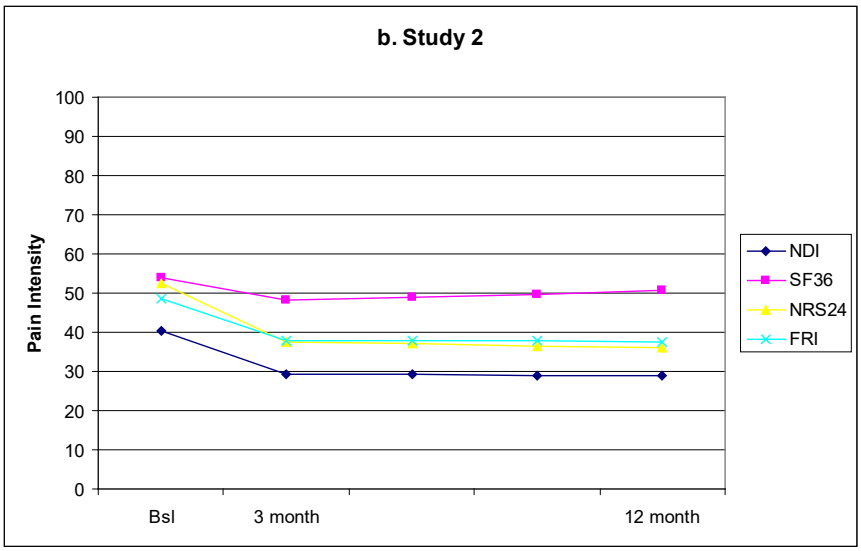
Mixed models regression was conducted within each cohort with subject as a random factor and recall period, response option and follow-up point as fixed factors, Table 4. As expected the effect of follow-up point was significant in all cohorts, that is pain scores were lower as time progressed after study inception regardless of how pain was measured. The effect of

recall period was comparable and significant in all cohorts, pain scores were generally higher when subjects were asked to provide an average rating over a longer time period than over a shorter or immediate time. The effects of number of response options was less clear. Study 1 contained two measures with 6-point verbal ratings scales and one 10cm VAS measure, in this cohort the VAS measure was associated with lower pain scores, when adjusted for recall time and follow-up point. Study 3 included two 6-point VRSs and three 11-point NRSs, here a small effect indicating lower pain scores for the scales with more response options was found. Study 2 included three 6-point VRSs and one 11-point NRS did not show an effect of the number of response options on pain score.

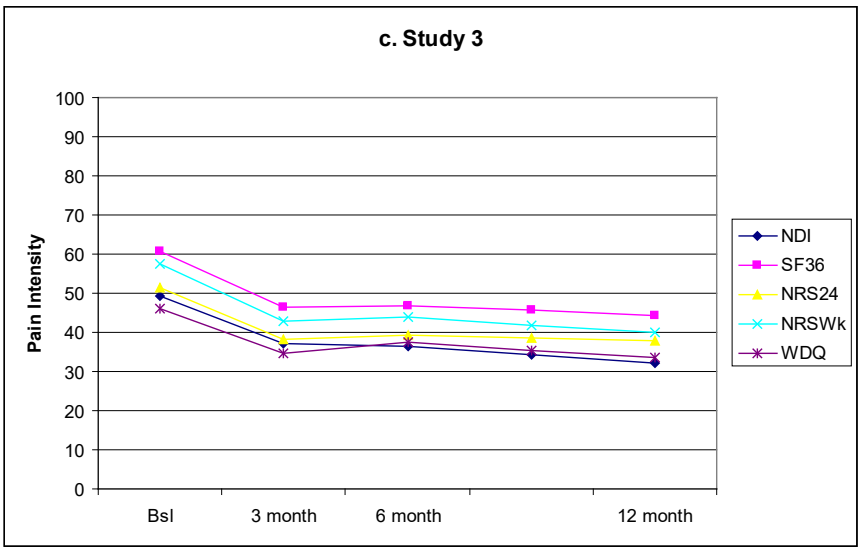
Fig. 1. Mean ratings (a) study 1 (NDI, SF36 and VAS). (b) study 2 (NDI, SF 36, FRI and NRS). (c) study 3 (NDI, SF 36, NRS 24/24, NRS Wk and WDQ).



182



183



184

185 **Table 3 – Mean scores and ANOVAs**

	Mean	SD	95% CI		n
			Lower	Upper	

Study 1						
	NDI	45.4	22.1	41.0	49.7	101
Baseline	SF 36	59.0	21.5	54.7	63.7	100
	VAS	36.4	22.9	31.9	41.0	100
ANOVA		F = 26.3 df = 2 p <0.01				
	NDI	28.1	24.4	23.1	33.2	91
3 Month	SF36	41.4	24.5	36.2	46.5	89
	VAS	20.9	24.2	15.4	26.4	78
ANOVA		F = 15.3 df = 2 p <0.01				
	NDI	24.7	25.4	19.2	30.1	86
6 Month	SF36	34.8	24.9	29.4	40.2	84
	VAS	19.8	21.1	15.1	24.4	82
ANOVA		F = 8.5 df = 2 p <0.01				
	NDI	25.6	27.0	19.9	31.3	89
12 Month	SF36	35.5	27.5	29.6	41.5	85
	VAS	23.0	26.0	16.8	29.2	70
ANOVA		F = 4.9 df = 2 p <0.01				
Study 2						
	NDI	40.5	17.9	37.4	43.5	134
Baseline	SF 36	53.9	20.5	50.3	55.6	125
	NRS24	52.6	20.0	49.2	56.0	134
	FRI	48.6	18.4	45.3	51.9	124
ANOVA		F = 13.1 df = 3 p <0.01				
	NDI	29.4	18.7	26.2	32.6	132
6 Week	SF36	48.2	21.9	44.2	52.1	120
	NRS24	37.4	24.1	33.2	41.5	132
	FRI	37.9	20.0	34.3	41.5	120
ANOVA		F = 16.4 df = 3 p <0.01				
	NDI	28.8	19.6	25.3	32.3	125
12 Month	SF36	50.6	22.8	46.2	54.9	110
	NRS24	36.2	24.6	31.8	40.5	125
	FRI	37.6	21.3	33.6	41.6	111
ANOVA		F = 19.3 df = 3 p <0.01				
Study 3						
	NDI	49.2	24.8	45.1	53.2	146
	SF36	60.8	16.3	58.2	63.5	146
Baseline	NRS24	51.5	20.4	48.1	54.8	146
	NRSWk	57.6	19.8	54.3	60.8	146
	WDQ	46.1	23.0	42.3	49.9	146
ANOVA		F = 12.1 df = 4 p <0.01				
	NDI	37.3	23.7	33.2	41.5	128
	SF36	46.3	21.2	42.6	50.0	127
3 Month	NRS24	38.3	22.7	34.3	42.3	127
	NRSWk	42.8	23.6	38.7	47.0	127
	WDQ	34.8	23.7	30.6	39.0	127
ANOVA		F = 5.1 df = 4 p <0.01				
	NDI	36.4	24.9	31.9	40.9	121
	SF36	46.8	22.8	42.7	51.0	120
6 Month	NRS24	39.3	26.1	34.5	44.0	120
	NRSWk	44.0	24.9	39.5	48.5	120
	WDQ	37.4	26.5	32.6	42.1	121
ANOVA		F = 3.9 df = 4 p <0.01				
	NDI	32.3	23.7	27.7	36.9	104
	SF36	44.4	24.3	39.7	49.2	104
12 Month	NRS24	37.9	25.2	32.9	42.8	103
	NRSWk	39.9	25.1	35.0	44.8	103
	WDQ	33.6	25.0	28.7	38.4	104
ANOVA		F = 4.1 df = 4 p <0.01				

186

Table 4 Mixed models regression

	Recall period (95% CI)	Response option (95% CI)	Follow-up point (95% CI)
Study 1	3.9 (3.0 to 4.8), p<0.01	-5.1 (-6.3 to -3.8), p<0.01	-6.5 (-7.6 to -5.5), p<0.01
Study 2	4.6 (3.9 to 5.3), p<0.01	0.58 (-1.2 to 2.4), p=0.53	-3.9 (-4.5 to -3.2), p<0.01
Study 3	4.0 (3.3 to 4.6), p<0.01	-2.0 (-3.4 to -0.7), p<0.01	-4.7 (-5.3 to -4.1), p<0.01

Estimates are unstandardised beta coefficients

Discussion

Main findings

As expected pain ratings from the different measures were strongly correlated, aside from a few low values, Pearson's r values were mostly within the range of 0.5 to 0.8, a finding that replicates previous studies[8]. The ANOVA analyses showed that there were significant differences between the pain scores extracted from the different measures, this was the case in all cohorts at all time-points. This provides evidence that patients score their pain intensity differently depending on how the question is framed and the score rated. Although intuitively sensible, empirical confirmation is important for both clinicians and researchers. The primary implication is the need for strict standardisation of pain assessment in both settings to ensure comparability of results.

Relationships and influences on measures

All pain scores reduced over the time course in the three cohorts and similar relationships were noticeable in both acute and chronic cohorts. Comparing the measures, the SF-36 score was consistently higher than the other pain ratings including the NDI, which also uses a 6-point VRS. The difference between SF-36 and NDI may have been due to the different descriptors used however we would contend that the difference in the length of time over which patients were asked to recall their pain is more likely responsible for the difference. We

base this contention on the fact that recall period had the most robust influence on pain score across all the studies and measures, even when adjusted for scale length and follow-up period. Some direct evidence in support of the influence of recall period also comes from comparison of the NRS scores in Study 3. At all time points scores for average pain over the past week were higher than those for average pain over the past 24 hours which in turn were higher than the question asking about pain today. Speculation that recall period influences symptom ratings has been raised previously[20, 21] and Broderick and colleagues report a similar influence in their series of studies[7, 22, 23]. Potential explanations include Ross's theory of implicit change[24], recall bias[25] and response shift[26]. The findings of this study fall broadly in line with previous work that examines the influence of recall time.

The importance of the number of scale items is less clear. While the results of the mixed models regression for Study 1 were suggestive of an influence, it is noted that the comparison in this study involved two 6-point VRSs and a 10cm VAS. As such it is difficult to determine whether the effect is due to the scale length or the choice of a discreet category versus marking a visual continuum. No significant effect of the number of response options was found in Study 2 and an effect of approximately half the size that for recall period was found in the Study 3 data. In their review of studies that compared different pain rating measures, Hjermstad et al[8] concluded that the number of response options is important. This conclusion was however limited to the assertion that while including more options potentially enables greater discrimination, relatively little is gained by having more than seven response options. On the basis of these data it seems unwise to draw firm conclusions as to the importance of scale length.

Limitations

We were unable to directly test the effect of different descriptors on the scales or the wording of the question. The VRS measures we had available (NDI, SF36 and FRI) are of comparable length but differ with respect to both recall period (immediate pain vs. 4-week average pain) and the wording of the descriptors, hence systematic scoring differences could be plausibly attributed to either of these reasons. The fact that all measures were not available in the three datasets reduces the power of our study, the fact that comparable effects for recall period were shown in the mixed models ANOVAs despite this adds credence to the findings. This study was not planned in advance of design and conduct of the source studies, as such it was a secondary analysis and findings should not be considered definitive.

Implications and future research directions

A number of implications can be drawn from this study. The first, that using different methods of measurement of the same construct yields systematically different results, is not controversial but important and reinforces the necessity of standardising measurement of pain in research and in the clinic. It lends weight to the call for cooperative moves towards deciding on common measures for research to be conducted in the future (The COMET initiative[27], IMMPACT[28]). The second implication is for researchers conducting meta-analyses, these results suggest that it may be unwise to rescale scores to a common metric by simple multiplication for the purposes of pooling. This is in line with the recommendations of the Cochrane Collaboration who suggest the use of standardised mean differences for estimating pooled treatment effects. It is noted however that the data in this study consists of raw scores over time, rather than between group differences, as such it is not clear that the same concerns apply. This question could be explored using between group difference data from RCTs.

Several questions also lead from this study. The first is to investigate the generalisability of the findings beyond those people with neck pain. While there appears little reason to believe that other clinical pain populations would respond differently validation of these findings in other samples would be useful, indeed the findings regarding recall period reinforce those by Broderick and colleagues in other populations[7, 22, 23]. A second question is to explore whether these findings hold for measures of other constructs, in particular patient reported disability and psychological function. In all cases such research would be ideally designed to test a-priori hypotheses about the function of different measures.

Conclusion

Measurement of pain in clinical populations is routinely performed in both the research and clinical environments, most commonly via single-item measures of pain intensity. Measurement instruments may differ in terms of the nature of the scale, the wording of the question, the recall period, the scale length and the descriptors on the scale all of which may potentially influence the score recorded by the patient. This study shows that commonly-used pain scales provided systematically different pain scores in the same patients. In particular asking patients to give an average pain rating over a longer period e.g. 4 weeks will yield a higher score than asking for an average over a shorter period e.g. 24 hours, or a rating of immediate pain.

Acknowledgements

Steve Kamper and Michele Sterling are funded by the National Health and Medical Research Council of Australia (NHMRC), Zoe Michaleff is funded by an Australian Postgraduate Award from the Australian government, Chris Maher is funded by the Australian Research Council (ARC). Funding for the studies was provided by the NHMRC, the ARC, the NSW

285 Motor Accidents Authority and the Motor Accidents Insurance Commission of Queensland.

286 None of the authors have conflicts of interest to declare.

287

References

[1] Litcher-Kelly L, Martino SA, Broderick JE, Stone AA. A systematic review of measures used to assess chronic musculoskeletal pain in clinical and randomized controlled clinical trials. Journal of Pain. 2007;8(12):906-13.

[2] Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs. 2005;14(7):798-804.

[3] Dijkers M. Comparing Quantification of Pain Severity by Verbal Rating and Numeric Rating Scales. J Spinal Cord Med. 2010;33(3):232-42.

[4] Vernon H. The Neck Disability Index: state-of-the-art. J Manipulative Physiol Ther. 2008;31:491-502.

[5] 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.

[6] Ostelo RWJG, De Vet HCW. Clinically important outcomes in low back pain. Best Prac Res Clin Rheumatol. 2005;19(593-607).

[7] Broderick J, Schwartz J, Schneider S. Can End-of-Day Reports Replace Momentary Assessment of Pain and Fatigue? J Pain. 2009;10(3):274-81.

[8] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, et al. Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review. J Pain Symptom Man. 2011;41(6):1073-93.

[9] Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain. 1986;27(1):117-26.

[10] Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. Pain. 1975.

312 [11] Kamper SJ, Rebbeck TJ, Maher CG, McAuley JH, Sterling M. Course and prognostic
313 factors of whiplash: a systematic review and meta-analysis. Pain. 2008 Apr 14;138(3):617-29.
314 [12] Spitzer WO, Skovron ML, Salmi LR, Cassidy JD, Duranceau J, Suissa S, et al.
315 Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders:
316 redefining "whiplash" and its management. Spine. 1995 Apr 15;20(8 Suppl):1S-73S.
317 [13] Walton DM, Pretty J, Macdermid JC, Teasell RW. Risk factors for persistent problems
318 following whiplash injury: Results of a systematic review and meta-analysis. J Orthop Sports
319 Phys Ther. 2009;39(5):334-50.
320 [14] TRACsa: Trauma and Injury Recovery SA. Clinical guidelines for best practice
321 management of acute and chronic whiplash associated disorders: Clinical resource guide.
322 Adelaide 2008.
323 [15] Kamper SJ, Maher CG, Hush JM, Pedler A, Sterling M. Relationship Between
324 Pressure Pain Thresholds and Pain Ratings in Patients With Whiplash-associated Disorders.
325 Clin J Pain. 2011;27:495-501.
326 [16] Stewart MJ, Maher CG, Refshauge KM, Herbert RD, Bogduk N, Nicholas M.
327 Randomized controlled trial of exercise for chronic whiplash-associated disorders. Pain.
328 2007;128(1-2):59-68.
329 [17] Michaleff ZA, Maher CG, Jull G, Latimer J, Connelly LB, Lin CWC, et al. A
330 randomised clinical trial of a comprehensive exercise program for chronic whiplash: trial
331 protocol. BMC Musculsoskeletal Dis. 2009;10:149.
332 [18] Peters ML, Patijn JP, Lamé I. Pain Assessment in Younger and Older Pain Patients:
333 Psychometric Properties and Patient Preference of Five Commonly Used Measures of Pain
334 Intensity. Pain Med. 2007;8(7):601-10.
335 [19] Willis C, Niere KR, Hoving JL, Green S, O'Leary EF, Buchbinder R. Reproducibility
336 and responsiveness of the Whiplash Disability Questionnaire. Pain. 2004;110(3):681-88.

337 [20] Kamper SJ, Maher CG, Mackay G. Global change rating scales; a review of strengths
338 and weaknesses and considerations for design. J Manual Manipulative Ther. 2009;17(3):163-
339 70.
340 [21] Kamper SJ, Ostelo RJW, Knol DL, Maher CG, De Vet HCW, Hancock MJ. Global
341 Perceived Effect scales provided reliable assessments of health transition in people with
342 musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin
343 Epidemiol. 2009;63(7):760-66.
344 [22] Broderick JE, Schneider S, Schwartz JE, Stone AA. Interference with activities due to
345 pain and fatigue: accuracy of ratings across different reporting periods. Qual Life Res.
346 2010;19(8):163-70.
347 [23] Broderick JE, Stone AA, Calvanese P, Schwartz JE, Turk DC. Recalled Pain Ratings:
348 A Complex and Poorly Defined Task. Journal of pain. 2006;7(2):142-49.
349 [24] Ross M. Relation of implicit theories to the construction of personal histories. Psychol
350 Rev. 1989;96(2):341-57.
351 [25] Herrmann D. Reporting current, past and changed health status. What we know about
352 distortion. Med Care. 1995;33(4):AS89-AS94.
353 [26] Schwartz CE SM. Meth odological approaches for assessing response shift in
354 longitudinal health-related quality-of-life research. Soc Sci Med. 1999;48:1531-48.
355 [27] http://www.comet-initiative.org/. [cited; Available from:
356 [28] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al.
357 Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain.
358 2005;113(1-2):9-19.
359
360