Original Article

THE BRITISH PAIN SOCIETY

British Journal of Pain
2016, Vol. 10(3) 141–155

DOI: 10.1177/2049463716642600
bjp.sagepub.com
SAGE

Magnitude and variability of effect sizes for the associations between chronic pain and cognitive test performances: a meta-analysis

Michél Rathbone[1], William Parkinson[2], Yasir Rehman[3], Shucui Jiang[4], Mohit Bhandari[5] and Dinesh Kumbhare[6]

Abstract

Objectives and Methods: A systematic review and meta-analysis were performed to estimate the size and variability of the association between chronic pain (CP) and poorer cognitive test performances as a function of individual tests, pain sub-types, and study sources on 22 studies having (1) a control group, (2) reported means and standard deviations (SDs) and (3) tests studied at least 3 times.

Results: CP patients performed significantly poorer with small to moderate effects (d = –.31 to –.57) on Digit Span Backward; STROOP Word; Color and Color-Word; Digit Symbol; Trail Making A and B; Rey Auditory Learning Immediate and Delayed Recall and Recognition. For these 10 measures, single effects (no interaction) were supported (I^2 = 0%–8%) and Random and Fixed models yielded similar results. No group differences were found for Corsi Blocks Forward or Wisconsin Cart Sorting Test Categories Achieved, or Perseveration. Effects for the Rey Complex Figure Immediate and Delayed Recall were significant, but effect size was inconclusive, given moderate to high heterogeneity and lack of consistency between Random and Fixed models. For the Paced Auditory Serial Addition Test, there was a homogeneous (I^2 = 0%) and significantly lower performance in fibromyalgia (d = –.47), but no effect in diagnostically undifferentiated pain samples, and wide variability across studies of whiplash (d = –.15 to –1.04, I^2 = 60%).

Conclusion: The magnitude and consistency of the CP – cognition effect depended on the test, pain subgroup and study source.

Summary points

1. Among tests showing a chronic pain (CP) – cognition effect, the magnitude of this association was consistently small to moderate across tests.
2. Effect size estimation was inconclusive for Digit Span Forwards, the Paced Auditory Serial Addition Test and the Rey Complex Figure Test.
3. Variance was too heterogeneous for testing cognitive domain specificity of the CP – cognition effect.

Keywords

Chronic pain, memory, cognitive testing, cognitive performances, meta-analysis

[1]Department of Medicine, McMaster University, Hamilton, ON, Canada
[2]School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada
[3]Health Sciences, McMaster University, Hamilton, ON, Canada
[4]Department of Biochemistry, McMaster University, Hamilton, ON, Canada
[5]Department of Surgery, McMaster University, Hamilton, ON, Canada
[6]Department of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding author:
William Parkinson, School of Rehabilitation Science, McMaster University, 1400 Main Street West, Hamilton, ON L8S 1C7, Canada.
Email: bparkins@mcmaster.ca

Introduction

Evidence for the presence of an association between chronic pain (CP) and poorer cognitive test performances has been summarized in systematic reviews[1–5] and previous meta-analyses of cognitive domains using test pooling.[1,2] Neurocognitive correlates of CP could add to disability[3] or illness role identification, increase physical or emotional distress, reduce motivation/self-efficacy, add to activity suppression by pain or increase pain coping errors. Altered cognitive processing may have common physiological denominators with CP.[5,6] Pain-associated cognitive impairment increases difficulties with the differential diagnosis between post-concussion syndrome related to brain trauma[7,8] and a post-concussion – like syndrome related to pain or non-brain injury;[9-12] a diagnostic problem made bigger by the high frequency of pain and head injury co-occurrence.[10,13,14]

Suggested mechanisms of this association include the following: shifting or individual differences in brain resource allocation,[15–18] pain-activated neuromodulators,[19] neuroplastic changes[20] or individual neurological susceptibilities.[21,22] Reduced grey matter has been found in CP patients[23–27] and an older literature focused on brain blood flow differences in CP.[28,29]

Some statistical concepts are of high relevance to our study. Meta-analysis involves increasing sample size by pooling subjects across studies to estimate the 'true' difference between population means. Measuring variance heterogeneity is critical to accepting, evaluating, and understanding the findings in meta-analyses.[30–32] Some study-to-study inconsistency is a natural result of sampling from truly different normal and CP distributions (i.e. sampling error). However, heterogeneity can also reflect: (1) an important interaction or moderator (i.e. between study variance is not '0') so that multiple true effect sizes apply or (2) excessive within study variance (or error) related to such variables as study differences in the specifics of test administration, symptom duration or other patient characteristics, the degree of subject matching in group selection, and mathematical control for covariates, making effect size estimation too inaccurate. As an example of the former, if neurocognitive effects are different for pain subgroups (e.g. fibromyalgia versus whiplash) heterogeneity could reflect multiple true effect sizes. This could also occur if different tests are combined to create a cognitive domain such as attention, or executive function. The I^2 statistic[30,31] is a measure of variance heterogeneity that controls for chance study-to-study differences and has been considered to minimize lowered or inflated heterogenieity by small or large numbers of studies. Finally, random effects models are generally preferred to fixed (or common) effects analysis because the former makes no assumption that the studies are drawing from a single super-population with one effect size. In the random model, variance is a combination of between and within study variance. However, the random model requires not only a sufficiently large number of subjects, but also a sufficiently large numbers of studies to accurately estimate the between study variance,[33] whereas the fixed effects model does not. Therefore, it should be instructive to examine statistical results using both models.

Collapsing across different cognitive tests thought to measure the same cognitive domain is common in meta-analyses of other clinical conditions.[20,34,35] However, cognitive tests are generally not specific to a cognitive domain. Cognitive domain analyses have yielded variable results. For example, in depression,[34] meta-analysis revealed no heterogeneity of variance in the 'attention' domain ($I^2 = 0\%$), moderate to high heterogeneity for 'working memory' ($I^2 = 61\%$) and 'psychomotor speed' ($I^2 = .68$), and very high variance heterogeneity for 'verbal' ($I^2 = 81\%$) and 'visual memory' ($I^2 = 88\%$).

Kessels et al.[4] provided a meta-analysis of 14 studies of the 'late whiplash syndrome' with emphasis on post-concussion complaints. They found poorer performance in whiplash versus controls with moderate to high effect sizes between −.53 and −.83 across cognitive domains, including attention, working memory, immediate and delayed verbal recall, visual scanning, and mental flexibility. Presence/absence of pain was not an independent variable in the analyses.

Berryman et al.[1] have provided meta-analyses on short-term memory and executive functions[2] in CP and have shown that CP is likely associated with lower performances, with small to moderate effect sizes in both domains. For short-term memory, meta-analysis showed lower scores in CP versus controls in 5 of 6 cognitive domains. Effect sizes ranged from −.31 to −.54. These researchers found good variance homogeneity for Immediate Visual Memory ($I^2 = 0\%$) and Non-verbal Working Memory ($I^2 = 18\%$) supporting the conclusion that a single true effect size had been estimated for those test clusters, which might reflect the effect size for those cognitive domains. However, homogeneity was weaker for Immediate Auditory Recall ($I^2 = 40\%$) and Verbal Working Memory ($I^2 = 40\%$) leaving it unclear as whether the effect sizes could vary importantly as a function of subject characteristics, or measurement error. For 'Running Memory', there was a large effect at −1.5 but also an I^2 of 86%. Only 5 of the 19 tests/subtests of short-term memory in their dataset had been studied more than twice which prevented estimating effect sizes for individual tests. In their study on executive functions, 3 domains of pooled

tests were created. For Response Inhibition, CP did not differ from controls on the number of correct items but CP patients had slower performances (d = −.31). CP had fewer correct responses in Set Shifting (d = −.25) and Complex Executive Functions (e.g. Wisconsin Card Sort d = −.49) with slower performances in these domains (d = −.57, d = −.34 respectively). For Executive Functions, variance was homogeneous for the Set Shifting domain whether number of correct responses ($I^2 = 0\%$) or response time ($I^2 = 0\%$) were considered. Weaker consistency was obtained for the effects in the Response Inhibition response times ($I^2 = 35\%$), and number correct in Complex Executive Functions was inconsistent ($I^2 = 73\%$).

To our knowledge, no published meta-analyses have used an approach that starts with estimation of the magnitude and consistency/inconsistency (i.e. variance homogeneity) of effects for individual tests in CP. None have examined delayed recall. Effect sizes and inconsistency for domains (pooled results across studies and different tests) outside of short-term memory and executive function (e.g. processing speed, attention) have not been reported. No previous studies have determined whether study-test clusters with consistent findings could reveal the relative domain specificity versus generalization across domains for cognitive associations with CP. Our aim was to provide a systematic review of the literature and quantitative analysis to estimate the effect sizes and effect consistencies/inconsistencies for individual neuropsychological tests, and for cognitive domains, including, processing speed, attention, memory, and executive functions. For domains showing variance outside acceptable sampling error, we also sought to explore pain subtype and study source as potential moderator variables.

Methods

Search and inclusion/exclusion criteria

A systematic review and meta-analysis were conducted according to the Cochrane (PRISMA) guidelines[36] using the McMaster OVID Medline, PsycINFO, and EMBASE databases. The search included peer-reviewed journal articles published between 1 January 1946 and 31 August 2015. The search words were broad and included 'chronic pain', 'fibromyalgia', 'complex regional pain syndromes', 'musculoskeletal pain', 'myofascial pain syndromes', 'whiplash associated disorder', 'cognitive functions', 'information-processing speed', 'distraction', 'cognitive interference', 'executive function', 'memory', 'recall', 'attention', 'working memory', 'reaction time', 'mental competency', 'psychomotor performance', 'learning', 'Wechsler Scales', 'neuropsychological tests', and 'neuropsychological measurement'. In addition, each test found in the first extraction was used as a key word for further searches. The search was limited to humans, English language, experimental designs, neuropsychological assessment and age 18 or older, and only studies with the following inclusion/exclusion criteria were retained for the analyses.

Inclusion criteria

1. Cognitive tests with known measurement properties. Studies of other outcomes only (e.g. electrophysiological measures) or tests used in research only without normative data or published standardized protocols were excluded.
2. CP patients symptomatic for at least 6 months and having no known brain injury or disease. Studies on pain patients selected based on the following characteristics were excluded: previous stroke, traumatic brain injury, dementia, neurodegenerative or neuroinflammatory diseases.
3. Studies on patients selected for malingering, or substance abuse were excluded.
4. A healthy control group selected on the basis of a history of the absence of pain.
5. Reported group means and standard deviations.
6. Cognitive tests/subtests studied at least 3 times, by different researchers or in different pain subgroups.

One investigator (Y.R.) conducted the initial search and selected 103 articles that potentially met criteria and required full reading. All 103 were read by the first investigator who excluded studies that did not meet criteria 2–6. Remaining reports were read independently by the first investigator and the psychologist to exclude studies not meeting criterion 1. Data were tabled in a standardized Excel sheet, and each group comparison was checked by both investigators to confirm accuracy of inclusion, and direction of the results of the individual studies on each cognitive measure.

Bias risk

Each included study was examined for risk of bias using the Newcastle – Ottawa Guidelines.[37] A study accumulated 1 star/point for each bias control feature up to a maximum 9. Scoring was performed independently by 2 co-authors (W.P., Y.R.) to ensure that study reports were clear enough to yield close agreement, and then discrepancies were discussed to produce the final summary.

Cognitive tests

A licensed psychologist and neuroscientist (W.P.) supervised test data extraction, and performed the data

summary, analyses and interpretations. Different versions have been created for some of the tests. Studies were not excluded for use of a different version of a test. Common versions are as follows.

Wechsler scales. In Digit Span, the respondent is asked to immediately recall progressively longer strings of digits in the same order (Forward subtest), or in reverse order (Backward subtest). These subtests are sensitive to problems with immediate memory span and, in the backward task, to mental manipulation and hold time, and therefore, are considered tests of aspects of verbal working memory. In Digit Symbol Coding, numbers 1–9 have unique symbols. Subjects have 2 minutes to print as many corresponding symbols as possible under a string of random numbers. The test requires processing speed but is sensitive to numerous types of impairment.

Paced auditory serial addition test. Random numbers from 1 to 9 are presented, typically by audiotape, and the respondent adds each number to the previous one, not the previous total. Presentation speed is progressively increased. The test requires processing speed, sustained and divided attention, working memory and arithmetic skills.

Corsi blocks. Equally spaced blocks appear on a computer screen and light up in random order. The respondent attempts to recall the sequence in the same, or reverse order. Difficulty increases with the number of blocks. The test requires attention, and immediate memory for order and spatial location. Demand on working memory is increased in the reverse condition.

Trail making test. The time required to connect dots, numbered and spread randomly over a page, is recorded (Trails A). Good performance requires processing speed and visual scanning. In the B subtest, the respondent must quickly alternate between numbers and letters in ascending numeric and alphabetical order. This subtest requires speed, and mental flexibility or attention set shifting.

Test of everyday attention. This test was developed to achieve ecological validity. Selective Attention involves timed searches for destinations on a map, and telephone numbers in a directory while ignoring distractions (other destinations, other numbers). In Sustained Attention, the respondent must find the winning lottery number within a string of numbers, or keep track of the floor they are on in an elevator using a series of tones. In Attention Switching, the examinee is mentally riding an elevator and must switch from counting floors upwards to counting downwards when the elevator changes direction. In Working Memory, the examinee must both select the correct elevator tone (auditory trial), or arrow (visual trial), while counting the tones, or arrows to identify floors as they change.

Rey auditory verbal learning test. A list of words is read and the respondent attempts to immediately recall as many as possible. The list is repeated multiple times to index learning speed. A novel list is inserted and recalled to create interference, followed by an attempt to recall the first list. The first list is recalled again after a delay. The test ends with a recognition trial for items on the first list interspersed with new words.

Rey complex figure test. The subject copies a two-dimensional abstract diagram and then reproduces it from memory at different delays. The test is sensitive to problems with visual–spatial constructional abilities, and visual memory.

STROOP test. Subjects read out loud the randomly ordered colour names red, green and blue printed in black, on white paper, completing as many words as possible in a fixed time (alternatively, the time to read all 100 words can be used). Then, the colours are read/reported for 100 series of 4 'x's', printed in red, green or blue. In the interference subtest, the words red, green and blue are printed with mismatching of colour to word (e.g. red typed in blue ink) and the colour must be read out loud. The test requires processing speed, selective attention and the executive ability to inhibit a habitual response (reading of words) to produce the correct one (colour).

Wisconsin card sort. Four key cards with geometric shapes of different number and colour are used along with 128 response cards. The goal is to recognize the current rule/concept (e.g. shape, colour, number) in as few trials as possible and maintain the rule/cognitive set ignoring the other 2 variables until the rule changes. The examiner provides feedback after each presentation. The rule is changed without warning and the subject must detect that and discover the new rule efficiently, using the positive and negative feedback on each trial. The test requires reasoning including an integration of the efficient use of feedback, concept recognition, impulse control and mental flexibility.

Statistical analyses

Meta-analysis was performed using Review Manager Software 5.2.[36] Pain subgroups were combined/collapsed unless variance heterogeneity prompted investigating interaction effects. Each test used in a study was entered separately (i.e. the 'test comparison' approach)

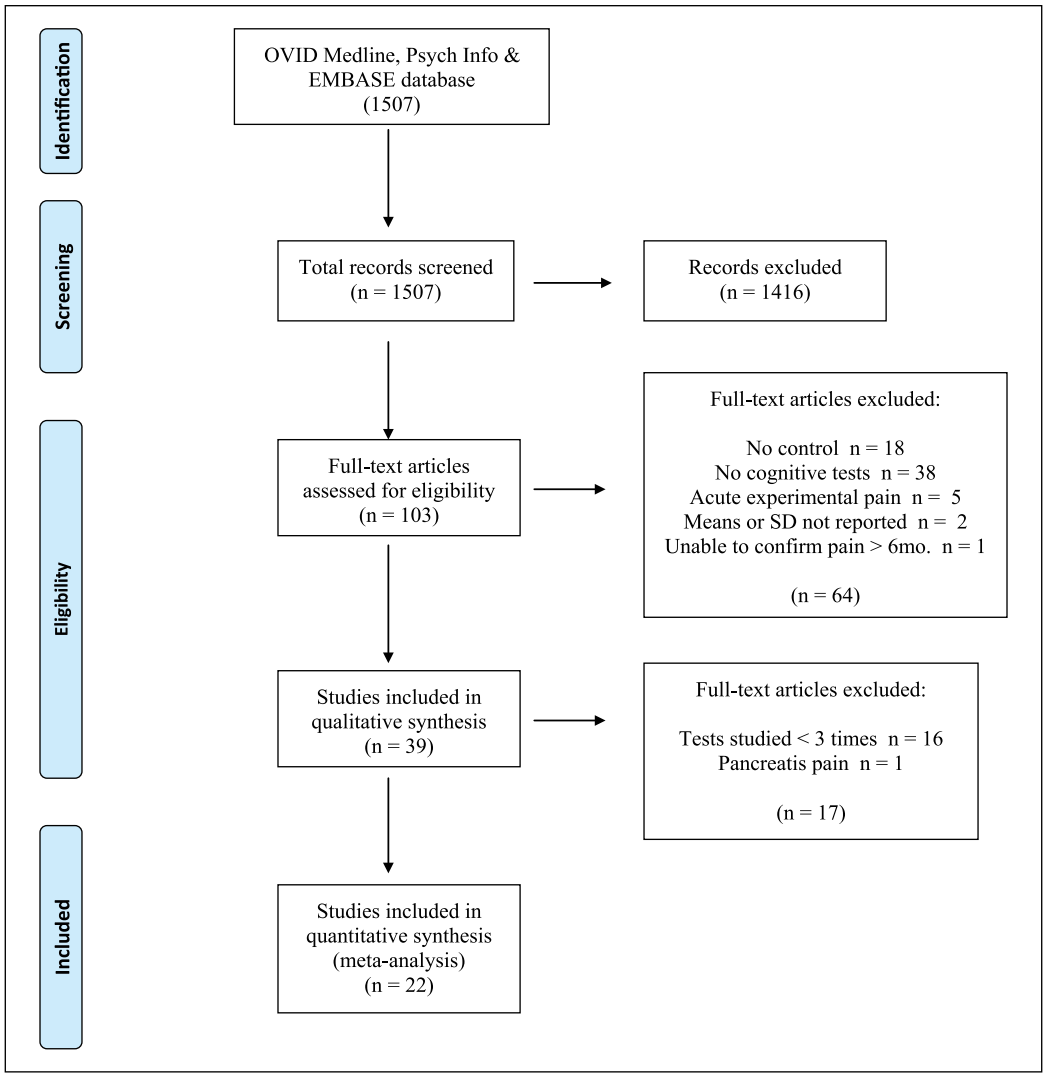

Figure 1. PRISMA 2009 flow diagram.
Reproduced from Moher D, Liberati A, Tetzlaff J, et al.; The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6(6): e1000097. DOI: 10.1371/journal.pmed1000097 with permission.[39]

and meta-analysis was performed on that test, across studies/published reports. The data were tabled so that weaker performances were consistently lower scores to adjust for study differences in how a test was administered and scored (e.g. stronger = lower time to complete a task, versus higher number of items in a fixed time). Effects were expressed as Standardized Mean Differences (Cohen's d). Statistical significance was set at alpha = .05 and extrapolated from the *Z* score. The 95% confidence intervals (CIs) were calculated for each meta-effect.

Variance heterogeneity was calculated for each meta-analysis using the I^2 statistic with I^2 30% or lower being acceptable, I^2 56% or higher being unacceptable[30,31] and intermediate I^2 suggesting caution relating to the potential for too much variance to estimate the effect, or a need for exploration of potential interaction effects (i.e. no single true effect). Random and Fixed Effects were both calculated to examine the consistency of the calculated effect sizes with and without consideration of between study variance. Effect estimates were interpreted to be small (0.2 standard deviation (SD)), moderate (0.50 SD), or large (0.80 SD) according to Cohen.[38]

In the event of moderate to high variance heterogeneity, meta-analyses were calculated for each pain subtype. Simple effects were calculated for single studies, when pooled effects for a pain subtype continued to yield moderate to high heterogeneity. Funnel plots were drawn by Review Manager as a check for potential publication bias.

Results

The systematic review yielded 1507 articles (Figure 1)[39]. Removal of duplicates, reviews, editorials and initial

Table 1. Newcastle – Ottawa Bias Risk Assessment.

Study	Selection	Comparability	Outcome	Total bias control points(x/9)
Andersson and Haldrup (2003)[40]	**	*	*	4
Bosma and Kessels (2002)[41]	**		*	3
Cánovas et al. (2009)[42]	****	*	*	6
Dick et al. (2002)[43]	****	*	*	6
Dick et al. (2008)[44]	****	*	*	6
Di Stefano and Radanov (1995)[45]	**	*	*	4
Gimse et al. (1997)[46]	***	*	*	5
Grace et al. (1999)[47]	****	*	*	6
Grisart and Van der Linden (2001)[48]	***		*	4
Kim et al. (2012)[49]	***	*	*	5
Landro et al. (1997)[50]	****		*	5
Lee et al. (2010)[51]	****	*	*	6
Oosterman et al. (2011)[52]	****	**	*	7
Oosterman et al. (2012)[53]	****	**	*	7
Roldan-Tapia et al. (2007)[54]	**		*	3
Schmand et al. (1998)[55]	***		*	4
Shur (2003)[56]	***		*	4
Sjogren et al. (2005)[57]	***		*	4
Verdejo-Garcia et al. (2009)[58]	***	*	*	5
Walitt et al. (2008)[59]	****		*	5
Walteros et al. (2001)[60]	**		*	3
Weiner et al. (2006)[61]	****		*	5

application of the inclusion and exclusion criteria left 103 papers that required full reading. Of these, 22 met criteria for the Main Analyses[40–61]. Among these, 21 tests/subtests had been studied at least 3 times. This permitted inclusion of Corsi Blocks Forward, but not Reverse. The 22 studies included 1193 participants.

Bias risk assessment ratings ranged from 3 to 7 out of 9 (Tables 1 and 2). All studies received 1 point on Exposure because the method of 'ascertainment' was cognitive testing for both groups. All studies lost 2 Outcome points, 1 for non-blinding whereby no authors described blinding examiners and it would be difficult to keep examiners blind to group status, and 1 for 'response rate', whereby no reports included a description of the proportions of identified subjects who participated. 'Selection' requirements were met variably with a range of 2–4 of 4 points. The main reason for loss of points was the potential for group selection bias. For Comparability, 1 of 2 maximum points was awarded if groups were equated on 1 or more tests estimating pre-morbid IQ ('most important factor') and a second point for 1 or more additional controlled variables. All studies included subject matching on age. Subject matching on gender and education was performed in 70% of studies, although both were not consistently controlled in the same studies. Subject matching on IQ was performed in less than ½ of studies. Only 1 group was able to show statistical similarity between groups on depression based on Minnesota Multiphasic Personality Inventory II Depression scale scores. Demonstrated control for subject effort or demand characteristics was also almost non-existent.

Meta-analyses on individual tests showed significantly poorer performances in CP over a wide range of cognitive tests (Table 3). Significant fixed effect sizes ranged between −.31 and −.57 standard deviations. The only larger effects were generated from studies all done by one team using the subscales of the Test of Everyday Attention. They found somewhat larger effect sizes ranging up to d = −.92 for the working memory subtest. Table 3 shows the initial summary of effect sizes for individual tests, based on the calculated overall effect, study-to-study consistency and the influence of between study variance.

Table 4 shows post-hoc analyses of tests with significant effects but moderate to high heterogeneity, in order to explore interactions. Tables 3 and 4 together showed the following overall results.

1. *No detectable effect*, based on effect magnitude and lack of statistical significance: Corsi Blocks Forward, Wisconsin Cart Sorting Test (WCST) Categories Achieved and Perseveration Responses.
2. *Significant, small to moderate effect*, based on effect magnitude, significant group differences, low I^2 and no difference between Random and Fixed models: Digit Span Backwards d = −.35; Trail Making A d = −.32 and B d = −.38; RAVLT Immediate d = −.52 and Delayed d = −.57

Table 2. Group Matching/Comparability.

Study	Age	Gender	Education	IQ	Depression	Effort/malingering	Socio-economic status
Andersson and Haldrup (2003)[40]	√	√					
Bosma and Kessels (2002)[41]	√	√	√		√		
Cánovas et al.(2009)[42]	√		√				
Dick et al. (2002)[43]	√	√	√				
Dick et al. (2008)[44]	√	√	√				
Di Stefano and Radanov (1995)[45]	√	√	√				
Gimse et al. (1997)[46]	√	√	√				√
Grace et al. (1999)[47]	√	√		√			
Grisart and Van der Linden (2001)[48]	√	√					√
Kim et al. (2012)[49]	√		√	√			
Landro et al. (1997)[50]	√	√	√				
Lee et al. (2010)[51]	√		√				
Oosterman et al. (2011)[52]	√	√		√			
Oosterman et al. (2012)[53]	√	√		√			
Roldan-Tapia et al. (2007)[54]	√			√			
Schmand (1998)[55]	√	√	√	√		√	
Shur (2003)[56]	√		√			√	
Sjogren et al. (2005)[57]	√	√	√				
Verdejo-Garcia et al. (2009)[58]	√	√	√				√
Walitt et al. (2008)[59]	√	√	√				√
Walteros et al. (2001)[60]	√			√			
Weiner et al. (2006)[61]	√	√	√				√

Recall and Recognition d = −.51; STROOP Word d = −.39, Color d = −37, and Color-Word d = −.35.

3. *Significant effect, magnitude inconsistent*: based on significant group differences, but moderate I^2: PASAT in fibromyalgia (d = −.57) and whiplash (d = −.59) but evidence against a more generalized effect in CP (d = −.15, n.s.) and evidence for either a diagnosis moderator effect, or potential excessive measurement error in whiplash.
4. *Inconclusive effect*, based on inconsistent effect size or statistical significance across Fixed and Random models, moderate to high I^2, and 95% CI close to '0': RCFT Immediate and Delayed Recall with both diagnosis and study source implicated as potential moderator variables and Digit Span Forward.

Test of Everyday Attention (TEA) subtests were excluded from the summary, and considered outliers requiring further research because they differ from other tests by larger effects, single laboratory study and conception (ecological focus).

Table 5 shows results for an exploratory analysis in which tests that individually showed low to moderate variance heterogeneity, were combined within models of cognitive domains. The RAVLT is the only verbal memory test that met our inclusion criteria, and therefore, the effect size is shown in Table 3, and it is not included in domain analyses. Insufficient numbers of tests studied 3 or more times with homogeneous variance prevented forming test clusters/domains emphasizing non-verbal attention, non-verbal processing speed, non-verbal working memory, verbal or non-verbal learning with immediate or delayed recall or for complex problem solving. It was possible to construct a domain emphasizing Executive Function – Simple Flexibility and Response Inhibition, and one emphasizing verbal attention, processing speed and, potentially, working memory. The findings show acceptable variance homogeneity and no differences between Random and Fixed models in these 2 test clusters. Figure 2 shows forest plots of the fixed effects outcomes for these 2 test clusters. The plots do not suggest consistent differences between tests within each domain on effect sizes or ranges.

Funnel plots (Figure 3) for domain analyses are largely symmetrical and show minimal departure of the highest weighted studies from the calculated effect sizes.

Discussion

Overall, the findings show poorer cognitive performances in CP patients with effect sizes ranging between small and moderate. Only 1 team using subscales of

Table 3. Results of meta-analyses.

Cognitive test	Study references	Total controls (N)	Total chronic pain (N)	Random effects	Random p <	Random I^2 (%)	Fixed effects	Fixed p <	Fixed I^2 (%)	Interpretation for effect size in CP
				Overall standardized mean difference (95% CI)			Overall standardized mean difference (95% CI)			
Digits Span Forward	Cánovas et al. 2009[42]-FM, Kim et al. 2012[49]-FM	87	93	–.33 (–.77, .08)	n.s.	46	–.31 (–.61, –.02)	0.05	46	Inconclusive
Digit Span Backward	Landro et al. 1997[50]-FM, Roldan-Tapia et al. 2007[54]-FM Roldan-Tapia et al. 2007[54]-RA	119	127	–.35 (–.61, –.10)	0.01	0	–.35 (–.68, –.10)	0.01	0	Small to moderate
Digit Symbol Coding	Grace et al. 1999[47]-FM, Lee et al. 2010[51]-UCP Schmand et al. 1998[55]-WL, Shur 2003[56]-FM Shur 2003[56]-UCP	1361	376	–.36 (–.5, –.22)	0.00001	8	–.34 (–.46, –.22)	0.00001	8	Small to moderate
TEA – Sustained	Dick et al. 2002[43]-FM, Dick et al. 2002[43]-UCP	90	90	–.77 (–1.07, –.46)	0.00001	1	–.77 (–1.07, –.46)	0.00001	1	Moderate to large
TEA – Selective	Dick et al. 2002[43]-RA, Dick et al. 2002[43]-UCP	90	90	–.87 (–1.18, –.56)	0.00001	0	–.87 (–1.18, –.56)	0.0001	0	Moderate to large
TEA – Switching	Dick et al. 2008[44]-FM	60	60	–.44 (–.8, –.07)	0.05	0	–.44 (–.80, –.07)	0.05	0	Small to moderate
TEA – Working Memory		90	90	–.95 (–1.33, –.56)	0.00001	33	–.92 (–1.24, –.64)	0.00001	33	Moderate to large
Trail Making – A	Bosma and Kessels 2002[41]-WL, Di Stefano and Radanov 1995[45]-6 months-WL	272	285	–.32 (–.49, –.15))	0.001	0	–.32 (–.49, –.15)	0.001	0	Small to moderate
Trail Making – B	Di Stefano and Radanov 1995[45]-24 months-WL, Gimse et al. 1997[46]-WL Oosterman et al. 2012[52]-UCP, Schmand et al. 1998[55]-WL Shur 2003[56]-FM, Shur 2003[56]-UCP Walitt et al. 2008[59]-FM, Walitt et al. 2008[59]-UCP Weiner et al. 2006[61]-TMT B only-UCP	432	448	–.38 (–.52, –.25)	0.00001	0	–.38 (–.52, –.25)	0.00001	0	Small to moderate
PASAT	Bosma and Kessels 2002[41]-WL, Di Stefano and Radanov 1995[45]-6 months-WL Di Stefano and Radanov 1995[45]-24 months-WL, Gimse et al. 1997[46]-WL Grace et al. 1999[47]-FM, Shur 2003[56]-FM Shur 2003[56]-UCP, Sjogren et al. 2005[57]-UCP	234	262	–.44 (–.69, –.19)	0.0005	43	–.39 (–.57, –.21)	0.0001	43	Significant Effect Magnitude Inconclusive
RAVLT – Immediate Recall	Schmand et al. 1998[55]-WL, Shur 2003[56]-FM, Shur 2003[56] UCP	168	186	–.52 (–.74, –.31)	0.00001	0	–.52 (–.74, –.31)	0.00001	0	Moderate
RAVLT – Delayed Recall	Kim et al. 2012[49]-FM-immediate and delayed only	112	133	–.57 (–.83, –.31)	0.0001	0	–.57 (–.83, –.31)	0.0001	0	Moderate

Table 3. (Continued)

Cognitive test	Study references	Total controls (N)	Total chronic pain (N)	Random effects	Random p <	Random I^2 (%)	Fixed effects	Fixed p <	Fixed I^2 (%)	Interpretation for effect size in CP
				Overall standardized mean difference (95% CI)			Overall standardized mean difference (95% CI)			
RAVLT – Recognition	Gimse et al. 1997[46]-WL, Grace et al. 1999[47]-FM-immediate only	88	110	–.51 (–.8, –.23)	0.001	0	–.51 (–.80, –.23)	0.001	0	Moderate
Corsi Blocks – Forward	Cánovas et al. 2009,[42] Di Stefano and Radanov 1995[45]-6 months-WL Di Stefano and Radanov 1995[45]-24 months-WL, Kim et al. 2012[49]	81	80	–.26 (–.57, .05)	n.s.	0	–.26 (–.57, .05)	n.s.	0	No Effect
RCFT – Immediate Recall	Bosma and Kessels 2002-WL,[41] Kim et al. 2012[49]-FM, Lee et al. 2010[51]-UCP	1357	350	–.59 (–1.11, –.07)	0.05	82	–.14 (–.26, .02)	0.05	82	Inconclusive
RCFT – Delayed Recall	Roldan-Tapia et al. 2007[54]-FM Roldan-Tapia et al. 2007[54]-RA Shur 2003[56]-FM, delayed only Shur 2003[56]-UCP, delayed only	1399	395	–.38 (–.68, –.09)	0.01	58	–.17 (–.28, –.05)	0.01	58	Inconclusive
Stroop – word reading speed	Oosterman et al. 2012[53]-UCP, Schmand et al. 1998[55]-WL	132	144	–.39 (–.64, –.14)	0.01	6	–.39 (.63, –.15)	0.001	6	Small to moderate
Stroop – colour reading speed	Walitt et al. 2008[59]-FM, Walitt et al. 2008[59]-UCP	132	144	–.37 (–.61, –.13)	0.01	0	–.37 (–.61, –.13)	0.01	0	Small to moderate
Stroop – coloured word reading speed		132	144	–.35 (–.54, –.15)	0.001	0	–.35 (–.59, –.11)	0.01	0	Small to moderate
Wisconsin Card Sort – Categories Achieved	Shur 2003[56]-FM	78	81	–.24 (–.77, .28)	n.s.	63	–.29 (–.60, .03)	P=0.07	63	Inconclusive
Wisconsin Card Sort – Perseveration	Shur 2003[56]-UCP Verdejo-Garcia et al. 2009-FM[58]	78	81	–0.15 (–.16, 0.47)	n.s.	0	–0.15 (–.47, –.16)	n.s.	0	No Effect

Test of Everyday Attention (TEA), Paced Auditory Serial Addition Test (PASAT), Rey Auditory Verbal Learning Test (RAVLT), Rey Complex Figure Test (RCFT). Pain subgroups include: fibromyalgia (FM), rheumatoid arthritis (RA), whiplash (WL), unspecified chronic pain (UCP) "Inconclusive" effect size estimation when variance heterogeneity moderate to high, also evident in 95% confidence interval, and in differences between Random and Fixed models.

Table 4. Examination of the influence of pain subgroup and study source on effect sizes.

Cognitive test	Sub-group	Study references	Total controls (N)	Total chronic pain (N)	Meta-random effects and simple effects standardized mean differences (95% CI)	p <	I^2 (%)	Interpretation of effect magnitude and heterogeneity in Table 3
Digit Span Forward	FM-Pooled	Kim et al. (2012),[49] Roldan-Tapia et al. (2007),[54] Cánovas et al. (2009),[42] Landro et al. (1997)[50]	78	72	–.16 (–.48, .16)	n.s.	0	No effect in FM
	Arthritis	Roldan-Tapia et al. (2007)[54]	15	15	–1.23 (–2.02, –.44)	0.01		Not sufficiently studied
PASAT	FM-Pooled	Shur (2003),[56] Grace et al. (1999) [47]	53	51	–.57 (–.97, –.18)	0.0001	0	Complex Findings:
	WL-Pooled		98	96	–.59 (–1.05, –.12)	0.01	60	Significant negative effect, inconsistent range small to large
	WL	Bosma and Kessels (2002)[41]	30	31	–.26 (–.76, .25)	n.s.		Diagnosis is a potential moderator
	WL	Di Stefano and Radanov (1995)[45]	42	42	–1.04 (–1.49, –.58)	0.000001		Variance created by study source suggests excessive measurement error could be operating for WL
	WL	Gimse et al. (1997)[46]	26	23	–.15 (–.71, .41)	n.s.		
	UCP-Pooled	Shur (2003),[56] Sjogren et al. (2005)[57]	85	113	–.15 (–.43, .13)	n.s.	0	
RCFT – Immediate Recall	FM-Pooled				–.87 (–1.65, –.09)	0.05	63	Inconsistent effect
	FM	Kim et al. (2012)[49]	24	23	–1.25 (–1.88, –.62)	0.0001		
	FM	Roldan-Tapia et al. (2007)[54]	15	15	–.45 (–1.17, –0.07)	0.05		
	Arthritis	Roldan-Tapia et al. (2007)[54]	15	15	–.74 (–1.48, .01)	0.05		
	WL	Bosma and Kessels (2002)[41]	30	31	–.7 (–1.2, –.18)	0.01		
	UCP	Lee et al. (2010)[51]	1273	266	–.02 (–.16, .11)	n.s.		
RCFT – Delayed Recall	FM-Pooled				–.58 (–1.13, –.02)	0.05	56	Inconsistent effect
	FM	Kim et al. (2012)[49]	24	23	–1.11 (–1.73, –.50)	0.001		
	FM	Roldan-Tapia et al. (2007)[54]	15	15	–.37 (–1.09, .35)	n.s.		
	FM	Shur (2003)[56]	21	23	–.23 (–.82, .36)	n.s.		
	Arthritis	Roldan-Tapia et al. (2007)[54]	15	15	–.66 (–1.4, .08)	0.08		
	WL	Bosma and Kessels (2002)[41]	30	31	–.49 (–1.0, .02)	0.06		
	UCP-Pooled	Lee et al. (2010),[51] Shur (2003)	1294	287	–.08 (–.21, .05)	n.s.	0	
Wisconsin Card Sort (WCST)	FM-Pooled	Shur (2003),[56] Verdejo-Garcia et al. (2009)[58]	57	59	–.22 (–1.1, .65)	n.s.	0	No Effect
Categories Achieved	UCP	Shur (2003)[56]	21	22	–.24 (–.85, .36)	n.s.		

Table 5. Pooled tests: cognitive domains.

Cognitive domain test clusters	Available test data for clustering	Effect	Total controls (n)	Total chronic pain (n)	Pooled standardized mean difference (95% CI)	p <	Pooled I^2 (%)	Domain cluster effect size in CP
1. Verbal Attention, Working Memory, Processing Speed	TMT-A, DSymbol, Stroop C, Stroop W, Digit Span Backwards	Random	2030	996	–.35 (–.43, –.27)	0.00001	0	Small to moderate
		Fixed	2030	996	–.35 (–.43, –.27)	0.00001	0	
2. Non-verbal Attention, Working Memory, Processing Speed	Insufficient variation of tests for converging evidence							Unknown
3. Verbal Learning, Recognition, Immediate and Delayed Recall	Insufficient variation of tests for converging evidence							Unknown
4. Non-verbal Learning, Recognition, Immediate and Delayed Recall	Heterogeneous variance for RCFT							Unknown
5. Executive Functions (Simple Flexibility and Response Inhibition)	TMT – B, Stroop Color – Word	Random	657	629	–.37 (–.48, –.26)	0.00001	0	Small to moderate
		Fixed	657	629	–.36 (–.46, –.26)	0.00001	0	
6. Executive Functions (Complex Problem Solving)	Insufficient variation of tests for converging evidence							Unknown

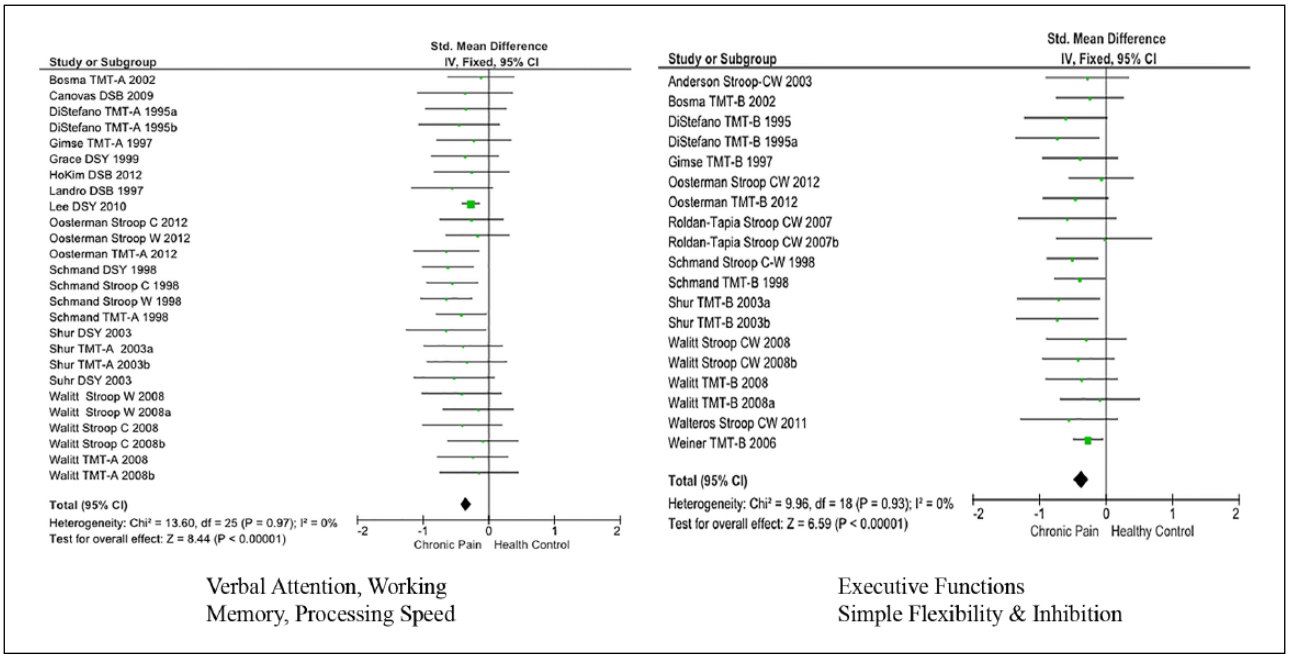

Figure 2. Forest plots: test clusters.

the TEA found consistently larger effect sizes. These remained under 1 standard deviation weaker performance in CP.

The higher effect sizes with the TEA could reflect measurement error. However, the TEA differs from other attention tests in that it was designed for ecological validity. This difference may mean that effect sizes for TEA subtests are larger than for other attention tests, but this possibility needs to be explored.

The present findings do not alter general conclusions of other authors.[1–5] The findings continue to show generally weaker cognitive performances in CP, weaker processing speed,[20] weaker performance on short-term memory measures,[1] weaker performances

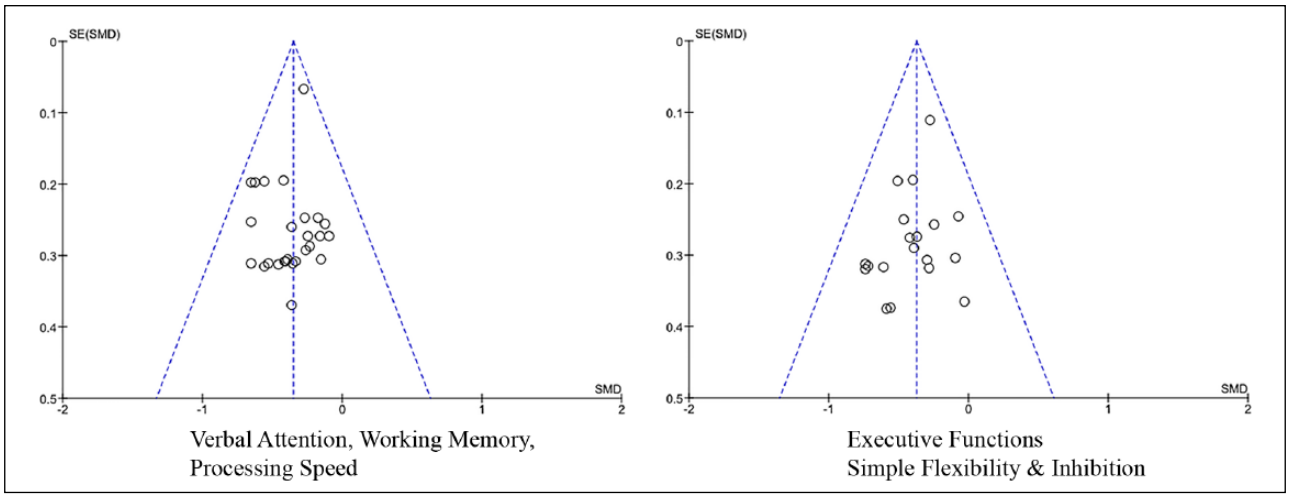

Figure 3. Funnel plots: test clusters.

on some executive functions[2] and no clear interaction between the presence of CP and multiple cognitive domains.[5] However, our emphasis on homogeneity measurement provides evidence that, within tests, detectable cognitive performance weakness, in the small to moderate range in CP patient samples, is a consistent finding for some tests, but not others. It was possible to demonstrate small to moderate effect sizes for Digit Span Backwards, Digit Symbol Coding, Trail Making A and B, and the 3 subtests of the STROOP, and moderate effect sizes for the 3 subtests of the Rey Auditory Verbal Learning Test. No effect appears to exist for Corsi Blocks Forward and Perseveration Responses or Categories Achieved on the WCST. Group differences could not be concluded or ruled out for Digit Span Forwards, or immediate or delayed recall on the Rey Complex Figure Test. There was a significantly reduced performance on the Paced Auditory Serial Addition Test in CP, but the effect magnitude varied considerably across studies and, therefore, remains unclear.

I^2 values can underestimate heterogeneity when combining small numbers of studies in meta-analyses,[62,63] and this has prompted a line of research to establish effective CIs for measures of variance heterogeneity. Our finding that I^2 was 0% for each of the 2 test clusters (Table 5), despite the use of many more studies than was the case for individual tests, increases confidence that variance heterogeneity was not missed in the analyses.

The source of the poorer cognitive performances in CP remains unclear. All studies equated groups on age. In all, 16 of the 22 deliberately equated on gender, and 16 equated on education. Other potentially confounding factors that were less frequently controlled include group differences in socioeconomic status, general intellect, medication use, mood and other co-morbid psychopathology, and effort and demand characteristics. Studies varied considerably on the extent to which groups were equated on each of these variables.

Another line of research has focussed on capturing differences in brain physiology, which could be relevant to determining whether pain per se negatively affects the brain's ability to perform cognitive tasks.[64,65] Most of the research involves static comparisons of CP patients with control samples. Longitudinal studies starting in acute pain or early following musculoskeletal injury are needed. Studies examining within-subject changes in cognitive performances to pain relieving interventions could be helpful. Nevertheless, intervention for pain could change a number of variables other than pain experience that could contribute to differences in brain physiology including stress levels, medication and other substance use, co-morbid psychopathology, activity levels and even conceivably, motivation.

The use of test clusters to investigate cognitive domains was possible only for an executive functioning domain that included tests sensitive to simple mental flexibility and response inhibition, and for a domain emphasizing verbal attention and processing speed with the possibility of working memory being relevant based on digit span backwards data. For both test clusters, CP performances were poorer than controls with the effects being small to moderate. Insufficient variation in tests studied multiple times and showing consistent results for each test prevented the creation of test clusters for non-verbal attention, working memory or processing speed, for verbal and non-verbal learning and memory, or for complex problem solving.

Our exploratory domain analysis did not reveal domain specificity of the association between CP and cognitive impairment. The fact that CP patients performed weaker on a wide range of tasks suggests that identifying domain specificity, if it exists, could be

difficult. The CP – cognitive impairment association may be highly domain non-specific. Despite the large number of studies over the past 20 years, research to date has not provided a sufficient number of repeatedly studied tests to provide large sample sizes and converging evidence (test clusters) to quantify the effect sizes for very many cognitive domain abilities in CP. The fact that the detectable effect sizes are largely well under 1 standard deviation forces a need for very large samples to investigate interaction effects.

Another factor contributing to difficulties testing domain specificity is that multiple domains can be variably relevant to each test. This has likely contributed to differences in domain definitions across study sources. Across meta-analysis studies in other health conditions, the structure of cognitive domains imposed on the data varies. For example, Digit Span forward and backward have been included together under Executive Function[20] or separated with digit span forward under Attention and backward under Working Memory.[34] Attention and working memory are sometimes combined as one construct.[35] Trail Making B was included under Executive Function in 1 study,[20] but under Attention Switching and excluded from the Cognitive Flexibility domain in another.[34] These study differences are understandable when tests are sensitive to multiple cognitive abilities or when terms such as 'attention', 'flexibility', and 'executive functions' are being used differently enough by different researchers. Our analyses were not meant to capture the best cognitive domain model for CP. Rather, we demonstrate first that variance homogeneity supports conducting at least some domain analyses in CP, but second that the effect sizes and consistency will depend on which tests are included in which domains.

It is noteworthy that the mix of studies differs between our analyses, and the 2 previous reports on meta-analyses in CP.[1,26] Those 2 studies included 17 reports that were excluded from our analyses. We excluded 14 because the tests had not been studied 3 or more times, one[66] because the patients had alcohol abuse related pancreatitis, one[67] because means and standard deviations were not available, and one because CP could not be confirmed.[68] In contrast, we included 8 studies excluded from their reports.[39,41,42,45,46,49,55,59] Six of these were studies of whiplash. Reasons for their exclusion of studies by Cánovas et al.[42] and Kim et al.[49] are not known. Whiplash studies may have been excluded given the possibility for brain trauma. We found good variance homogeneity for the Digit Symbol Coding, Trail Making A & B, the RAVLT, and the STROOP despite inclusion of whiplash samples. It is possible that whiplash contributed to high variance across studies of the PASAT. More generally, the similar findings between the present analyses and the former reports provides evidence that weaker cognitive performance in CP is robust and characteristic of a wide range of tests and likely multiple cognitive domains. We provide evidence that some tests appear to be less implicated in CP than others. Given that some tests included in our analyses showed inconsistent findings across studies, and given the large number of tests that have not been repeatedly researched in CP, much remains unknown about the association between CP and cognitive functioning.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was self-funded by the authors.

References

1. Berryman C, Stanton TR, Bowering J, et al. Evidence for working memory deficits in chronic pain: a systematic review and meta-analysis. *Pain* 2013; 154: 1181–1196.
2. Berryman C, Stanton TR, Bowering KJ, et al. Do people with chronic pain have impaired executive function? A meta-analytical review. *Clin Psychol Rev* 2014; 34: 563–579.
3. Hart RP, Martelli MF and Zasler ND. Chronic pain and neuropsychological functioning. *Neuropsychol Rev* 2000; 10(3): 131–149.
4. Kessels RPC, Aleman A, Verhagen WIM, et al. Cognitive functioning after whiplash injury: a meta-analysis. *J Int Neuropsychol Soc* 2000; 6: 271–278.
5. Moriarty O, McGuire BE and Finn DP. The effect of pain on cognitive function: a review of clinical and preclinical research. *Prog Neurobiol* 2011; 93: 385–404.
6. Chapman CR. Limbic processes and the affective dimension of pain. *Prog Brain Res* 1996; 110: 63–81.
7. Karzmark P, Hall K and Englander J. Late-onset postconcussion symptoms after mild brain injury: the role of premorbid, injury-related, environmental and personality factors. *Brain Inj* 1995; 9: 21–26.
8. Emanuelson I, Andersson Holmkvist E, Björklund R, et al. Quality of life and post-concussion symptoms in adults after mild traumatic brain injury: a population-based study in western Sweden. *Acta Neurol Scand* 2003; 108(5): 332–338.
9. Block C and Cianfrini L. Neuropsychological and neuroanatomical sequelae of chronic non-malignant pain and opioid analgesia. *NeuroRehabilitation* 2013; 33(2): 343–366.
10. Iverson GL and McCracken LM. 'Postconcussive' symptoms in persons with chronic pain. *Brain Inj* 1997; 11(11): 783–790.
11. Kewman DG, Vaishampayan N, Zald D, et al. Cognitive impairment in musculoskeletal pain patients. *Int J Psychiatry Med* 1991; 21: 253–262.

12. Radanov BP, Di Stefano G, Schnidrig A, et al. Cognitive functioning after common whiplash: a controlled follow-up study. *Arch Neurol* 1993; 50: 87–91.
13. Beaupré M, De Guise E and McKerral M. The association between pain-related variables, emotional factors, and attentional functioning following mild traumatic brain injury. *Rehabil Res Pract* 2012; 2012: Article ID 924692 (10 pp.).
14. Clark ME, Bair MJ, Buckenmaier CC 3rd, et al. Pain and combat injuries in soldiers returning from Operations Enduring Freedom and Iraqi Freedom: implications for research and practice. *J Rehabil Res Dev* 2007; 44(2): 179–194.
15. Crombez G, Van Damme S and Eccleston C. Hypervigilance to pain: an experimental and clinical analysis. *Pain* 2005; 116(1–2): 4–7.
16. Eccleston C. Chronic pain and distraction: an experimental investigation into the role of sustained and shifting attention in the processing of chronic persistent pain. *Behav Res Ther* 1995; 33: 391–405.
17. Eccleston C and Crombez G. Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol Bull* 1999; 125: 356–366.
18. Khatibi A, Dehghani M, Sharpe L, et al. Selective attention towards painful faces among chronic pain patients: evidence from a modified version of the dot-probe. *Pain* 2009; 142(1–2): 42–47.
19. DeQuervin DJ, Roozendaal B and McGraugh JL. Stress and glucocorticoids impair retrieval of long-term spatial memory. *Nature* 1998; 394: 787–790.
20. McDermott LM and Ebmeier KP. A meta-analysis of depression severity and cognitive function. *J Affect Disord* 2009; 19: 1–8.
21. Duric V and McCarson KE. Persistent pain produces stress-like alterations in hippocampal neurogenesis and gene expression. *J Pain* 2006; 7: 544–555.
22. Hu Y, Yang J, Hu Y, et al. Amitriptyline rather than lornoxicam ameliorates neuropathic pain-induced deficits in abilities of spatial learning and memory. *Eur J Anaesthesiol* 2010; 27(2): 162–168.
23. Apkarian AV, Hashmi JA and Baliki MN. Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. *Pain* 2011; 152: S49–S64.
24. Schmidt-Wilcke T, Leinisch E, Straube A, et al. Gray matter decrease in patients with chronic tension type headache. *Neurology* 2005; 65: 1483–1486.
25. Schmidt-Wilcke T, Leinisch E, Ganssbauer S, et al. Affective components and intensity of pain correlate with structural differences in gray matter in chronic back pain patients. *Pain* 2006; 125(1–2): 89–97.
26. Schmidt-Wilcke T, Luerding R, Weigand T, et al. Striatal grey matter increase in patients suffering from fibromyalgia – a voxel-based morphometry study. *Pain* 2007; 132(Suppl. 1): S109–S116.
27. Valfre W, Rainero I, Bergui M, et al. Voxel-based morphometry reveals gray matter abnormalities in migraine. *Headache* 2008; 48: 109–117.
28. Calandre EP, Bembibre J, Arnedo ML, et al. Cognitive disturbances and regional cerebral blood flow abnormalities in migraine patients: their relationship with the clinical manifestations of the illness. *Cephalalgia* 2002; 22: 291–302.
29. Martelli MF, Grayson R and Zasler ND. Post traumatic headache: psychological and neuropsychological issues in assessment and treatment. *J Head Trauma Rehabil* 1999; 14(1): 49–69.
30. Higgins JPT and Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539–1558.
31. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557–560.
32. Petitti DB. Approaches to heterogeneity in meta-analysis. *Stat Med* 2001; 20: 3625–3633.
33. Field AP. Is the meta-analysis of correlation coefficients accurate when population effect sizes vary? *Psychol Meth* 2005; 10: 444–467.
34. Lee RSC, Hermens DF, Melanie A, et al. A meta-analysis of cognitive deficits in first-episode Major Depressive Disorder. *J Affect Disord* 2012; 140: 113–124.
35. Scott JC, Matt GE, Wrocklage KM, et al. A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychol Bull* 141: 105–140.
36. The Cochrane Collaboration. *Review Manager (RevMan)* (Computer program), version 5.2. Copenhagen: The Nordic Cochrane Centre, 2014.
37. Wells GA, Shea B, O'Conell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses, 2009, http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm
38. Cohen J. Set correlation and contingency tables. *Appl Psychol Meas* 1988; 12(4): 425–434.
39. Moher D, Liberati A, Tetzlaff J, et al.; The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 6(6): e1000097. DOI: 10.1371/journal.pmed1000097
40. Andersson G and Haldrup D. Personalized pain words and Stroop interference in chronic pain patients. *Eur J Pain* 2003; 7: 431–438.
41. Bosma FK and Kessels RP. Cognitive impairments, psychological dysfunction, and coping styles in patients with chronic whiplash syndrome. *Neuropsychiatry Neuropsychol Behav Neurol* 2002; 15: 56–65.
42. Cánovas R, León I, Roldan MD, et al. Virtual reality tasks disclose spatial memory alterations in fibromyalgia. *Rheumatology* 2009; 48: 1273–1278.
43. Dick B, Eccleston C and Crombez G. Attentional functioning in fibromyalgia, rheumatoid arthritis, and musculoskeletal pain patients. *Arthritis Rheum* 2002; 47(6): 639–644.
44. Dick BD, Verrier MJ, Harken KT, et al. Disruption of cognitive function in fibromyalgia syndrome. *Pain* 2008; 139: 610–616.
45. Di Stefano G and Radanov BP. Course of attention and memory after common whiplash: a two-year prospective study with age, education and gender pair-matched patients. *Acta Neurol Scand* 1995; 91: 346–352.
46. Gimse R, Bjorgen IA, Tjell C, et al. Reduced cognitive functions in a group of whiplash patients with

demonstrated disturbances in the posture control system. *J Clin Exp Neuropsychol* 1997; 19(6): 838–849.
47. Grace GM, Nielson WR, Hopkins M, et al. Concentration and memory deficits in patients with fibromyalgia syndrome. *J Clin Exp Neuropsychol* 1999; 21: 477–487.
48. Grisart JM and Van der Linden M. Conscious and automatic uses of memory in chronic pain patients. *Pain* 2001; 94: 305–313.
49. Kim S-H, Kim S-H, Kim S-K, et al. Spatial versus verbal memory impairments in patients with fibromyalgia. *Rheumatol Int* 2012; 32: 1135–1142.
50. Landro NI, Stiles TC and Sletvold H. Memory functioning in patients with primary fibromyalgia and major depression and healthy controls. *J Psychosom Res* 1997; 42: 297–306.
51. Lee DM, Pendleton N, Tajar A, et al. Chronic widespread pain is associated with slower cognitive processing speed in middle-aged and older European men. *Pain* 2010; 51: 30–36.
52. Oosterman JM, Derksen LC, van Wijck AJM, et al. Memory functions in chronic pain: examining contributions of attention and age to test performance. *Clin J Pain* 2011; 27: 70–75.
53. Oosterman JM, Derksen LC, van Wijck AJM, et al. Executive and attentional functions in chronic pain: does performance decrease with increasing task load? *Pain Res Manag* 2012; 17(3): 159–164.
54. Roldán-Tapia L, Cánovas-López R, Cimadevilla J, et al. Cognition and perception deficits in fibromyalgia and rheumatoid arthritis. *Reumatol Clin* 2007; 3: 101–109.
55. Schmand B, Lindeboom J, Schagen S, et al. Cognitive complaints in patients after whiplash injury: the impact of malingering. *J Neurol Neurosurg Psychiatr* 1998; 64: 339–343.
56. Shur JA. Neuropsychological impairment in fibromyalgia: relation to depression, fatigue, and pain. *J Psychosom Res* 2003; 55: 321–329.
57. Sjogren P, Christrup LL, Petersen MA, et al. Neuropsychological assessment of chronic non-malignant pain patients treated in a multidisciplinary pain centre. *Eur J Pain* 2005; 9: 453–462.
58. Verdejo-Garcia A, Lopez-Torrecillas F, Calandre EP, et al. Executive function and decision-making in women with fibromyalgia. *Arch Clin Neuropsychol* 2009; 24: 113–122.
59. Walitt B, Roebuck-Spencer T, Bleiberg J, et al. Automated neuropsychiatric measurements of information processing in fibromyalgia. *Rheumatol Int* 2008; 28: 561–566.
60. Walteros C, Sánchez-Navarro JP, Muñoz MA, et al. Altered associative learning and emotional decision making in fibromyalgia. *J Psychosom Res* 2011;70: 294–301.
61. Weiner DK, Rudy TE, Morrow L, et al. The relationship between pain, neuropsychological performance, and physical function in community-dwelling older adults with chronic low back pain. *Pain Med* 2006; 7: 60–70.
62. Takkouche B, Khudyadov P, Costa-Bouzas J, et al. Confidence intervals for heterogeneity measures in meta-analysis. *Am J Epidemiol* 2013; 178(6): 993–1004.
63. Thorland K, Imberger G, Johnston BC, et al. Evaluation of heterogeneity estimates and their 95% confidence intervals in large meta-analyses. *PLoS ONE* 2012; 7(7): e39471–e39478.
64. Browning M, Fletcher P and Sharpe M. Can neuroimaging help us to understand and classify somatoform disorders? A systematic and critical review. *Psychosom Med* 2011; 73: 173–184.
65. Legrain V, Damme SV, Eccleston C, et al. A neurocognitive model of attention to pain: behavioral and neuroimaging evidence. *Pain* 2009; 144: 230–232.
66. Jongsma MLA, Postma SAE, Souren P, et al. Neurodegenerative properties of chronic pain: cognitive decline in patients with chronic pancreatitis. *PLoS ONE* 2011; 6: e23363.
67. Leurding R, Weigand T, Bogdahn U, et al. Working memory performance is correlated with local brain morphology in the medial frontal and anterior cingulated cortex in fibromyalgia patients: structural correlates of pain-cognition interaction. *Brain* 2008; 131: 3222–3231.
68. Melkumova KA, Podchufarova EV and Yakhno NN. Characteristics of cognitive functions in patients with chronic spinal pain. *Neurosci Behav Physiol* 2011; 41: 20–24.