The pain drawing as an instrument for identifying cervical spine nerve involvement in chronic whiplash-associated disorders

Gabriella Bernhoff, Maria Landén Ludvigsson, Gunnel Peterson, Bo Christer Bertilson, Madeleine Elf, Anneli Peolsson

Objective: The aim of the study was to investigate the psychometric properties of a standardized assessment of pain drawing with regard to clinical signs of cervical spine nerve root involvement.

Design: This cross-sectional study included data collected in a randomized controlled study.

Patients: Two hundred and sixteen patients with chronic (≥6 months) whiplash-associated disorders, grade 2 or 3, were included in this study.

Methods: The validity, sensitivity, and specificity of a standardized pain drawing assessment for determining nerve root involvement were analyzed, compared to the clinical assessment. In addition, we analyzed the interrater reliability with 50 pain drawings.

Results: Agreement was poor between the standardized pain drawing assessment and the clinical assessment (kappa = 0.11, 95% CI: −0.03 to 0.20). Sensitivity was high (93%), but specificity was low (19%). Interrater reliability was good (kappa = 0.64, 95% CI: 0.53 to 0.76).

Conclusion: The standardized pain drawing assessment of nerve root involvement in chronic whiplash-associated disorders was not in agreement with the clinical assessment. Further research is warranted to optimize the utilization of a pain/discomfort drawing as a supportive instrument for identifying nerve involvement in cervical spinal injuries.

Keywords: pain drawing, cervical vertebrae, diagnostic self-evaluation, radiculopathy, reproducibility of results, whiplash injuries

Introduction

A pain drawing is a low-cost tool for recording and assessing the location and distribution of pain. Pain drawings have been in use for >50 years. However, the utilization and modes of assessing pain drawings are not standardized in the clinic.

Allowing a patient to visualize the distribution of pain on a pain drawing may be advantageous. The pain sensation may be hard to explain verbally. Also, the clinical history relies on the patient’s subjective report. Indeed, it has been suggested that changes in the patient’s attention might influence reliability and that reporting pain with a pain drawing may favor reliability because it may be more focused than an oral description.

Pain drawing as an instrument has been studied with the patient’s drawing foremost not only as a simple descriptor, or an indicator of pathogenesis such as nerve involvement and psychosocial pain components, but also as a predictor of treatment outcome. Some previous studies focused on the aspects of the quality of this assessment in terms of validity, and most assessed low back pain. However, those studies have not provided a consensus on the most appropriate method of assessing a pain
drawing. To our knowledge, only one study\(^1\) investigated the validity of pain drawings for assessing neck discomfort; they concluded that pain drawings had a high sensitivity (90\%) for identifying neuropathic pain. Specificity was not analyzed because, in the final clinical assessment, most patients (92\%) were diagnosed with nerve involvement. According to our search in electronic databases, no studies have investigated whether pain drawings agreed with clinical criteria for determining nerve root involvement among individuals with chronic whiplash-associated disorders (WAD). The incidence of whiplash injury is estimated to be three of 1,000 inhabitants per year in Sweden\(^10\) and in Western Europe and North America.\(^1^1\) About half of the patients with WAD have persistent pain and disability.\(^1^2\) Symptoms include pain in the neck, which radiates to the arms, due to an unphysiological load that caused damage to neck tissues.\(^1^3,1^4\)

It has been suggested that cervical spine nerve involvement is an underestimated cause of pain or discomfort\(^1^5,1^6\) because it may not be detected on magnetic resonance imaging scans.\(^5,1^7\) This is likely to be the case when pain is mediated by chemical irritation.\(^1^9\) In this study, we assumed that dermatomes could provide sufficient guidance in pain topography. This guidance could enable identification of spinal segments that might be affected. Nerve involvement can manifest clinically as pain and paresthesia\(^1^9\) and/or loss of sensation or motor function.\(^2^0\) Neuropathic pain is diagnosed with a step-by-step procedure, where three criteria must be present: first, a history that suggests a relevant lesion or disease; second, a neuroanatomically plausible distribution; and third, confirmatory diagnostic tests.\(^2^1\) Pain due to nerve root involvement may be experienced within part of a dermatome but not necessarily within the entire dermatome.\(^2^1,2^2\) The use of dermatomes in diagnostics must be approximate due to overlapping regions and interindividual variabilities.\(^2^3-2^6\)

Due to the lack of knowledge on the diagnostic accuracy of pain drawings with regard to cervical spine involvement, this study aimed to investigate the criterion validity, sensitivity, and specificity of a standardized pain drawing for assessing clinical signs of nerve root involvement in chronic WAD. Another aim was to investigate the interrater agreement of the assessment of the pain drawings for nerve root involvement.

**Methods**

This study comprised a cross-sectional analysis of data retrieved from a randomized controlled trial (RCT).\(^2^7,2^8\) The RCT investigated three different physiotherapy interventions for patients with chronic WAD, and they conducted one follow-up at 12 months (in total, four follow-ups). Data on the WAD grades and pain drawings were recorded at the same consultation; however, the pain drawing was recorded at the patient at home, prior to the visit. This study consisted of two parts. In the first part, we used baseline data \(n=213\) to investigate the criterion validity, sensitivity, and specificity of pain drawings compared to a clinical examination. In the second part, we investigated the interrater reliability of three assessors of pain drawings \(n=50\) drawings.

**Participants**

Of 216 patients included, 142 women (65\%) and 74 men (35\%), 213 completed the pain drawings at baseline. The mean age was 40 years (SD: 11.4 years, range 18–63 years).\(^2^8\) The mean duration of WAD was 20 months (SD: 9.2 months, range 6–36 months). At baseline, patients evaluated pain on a visual analog scale (VAS); the mean score was 42 (SD: 24.3, range 20–62). The mean neck disability index\(^2^9\) score was 17 (SD 6.6, range 12–21). All data were normally distributed.

**Recruitment**

The same inclusion criteria used in the prior RCT reported by Ludvigsson et al\(^2^7\) were also used in this study. Inclusion criteria were age 18–63 years, a whiplash injury in the preceding 6–36 months, WAD grade 2 or 3, neck disability index \(\geq20\)\%, and an average VAS pain score \(\geq20\) mm. Exclusion criteria were known or suspected serious disease; prior fracture or luxation of the cervical spine; prior neck trauma with persisting symptoms; neck surgery; neck pain resulting in \(\geq1\) month sick leave during the year preceding the whiplash injury; signs of traumatic brain injury connected to the whiplash injury; generalized pain, with the main pain area at a location other than the neck; diseases or injuries that might prevent full participation in the study; serious psychological illness; known drug abuse; and Swedish language insufficiency (unable to fill out the questionnaire).

The clinical physical examination was conducted by experienced physiotherapists (mean, 18 years of experience). Test leaders underwent joint practice sessions before the start of recruitment to ensure that they conducted the clinical tests according to a standardized protocol. Details on the whiplash injury were recorded when performing the patient history; in cases of uncertainty, medical records were checked.

**Ethical aspects**

Ethical principles were followed in accordance with the Declaration of Helsinki. Patient data were recorded anonymously. The study was free from commercial interests and bindings. The RCT protocol was approved by the Regional...
Ethics Review Board in Linköping. All study participants provided informed consent.

**Diagnostic accuracy**

To examine whether pain drawings (n=213) could be used to identify nerve involvement in the cervical spine, the outcome of a standardized assessment was compared to the reference clinical WAD grade. WAD 2 was defined as a neck complaint and musculoskeletal signs, including decreased range of motion and point tenderness. WAD 3 was defined, according to Spitzer et al., as a neck complaint and neurological signs. WAD 3 was further defined, by Ludvigsson et al., as follows: 1) arm pain reported on a VAS or pain drawing and/or numbness/pain/prickling in arm(s) reported on the questionnaire, with no other known causes of arm symptoms, and 2) two or more neurological findings in the same dermatome and/or myotome with the following tests: sensitivity, muscle strength, reflexes of the upper extremity, and familiar arm pain provoked by nerve tension through traction/compression of the neck. In cases with arm pain, manual neck traction of the corresponding spinal segments (patient in supine position) was performed, where the pain had to be alterable in order for the condition to be assessed as WAD 3. The clinical assessments included both arms or body halves.

Each drawing was classified in the standardized assessment as nonnerve or nerve involvement. Only an assessment for the upper body half was performed because the clinical assessment focused exclusively on neck-related disorders. However, the lower extremity reflexes were tested to assess and exclude possible myelopathy. The pain drawing was a silhouette of the human body (Figure 1). The questionnaire instructed the study participants to: “Please shade the figure in all areas where you usually experience pain!”.

In this study, one experienced physician and two experienced physiotherapists assessed the pain drawings. The assessors were blinded, that is, they had no contact with study participants, the only available data were the pain drawings, and they were not involved in other aspects of the RCT.

For assessing pain drawings, a dermatome template was provided in a simple, two-dimensional layout. The dermatome template consisted of a transparent plastic sheet with the same human figure outline used in the questionnaire. The human segmental dermatomes were marked on separate frontal and dorsal views (Figure 1), as described by Lee et al. In the standardized assessment, the dermatome template was placed over the drawing from the questionnaire. A dermatome was recorded when the patient marked an area that corresponded to an area specified in the brief assessment guide.

The brief assessment guide included the following instructions. Each dermatome was to be screened in its entirety, on ventral and dorsal surfaces. For an affirmation of nerve involvement, at least one dermatome had to contain either two or more separate markings or one marking that followed the principal distribution of a dermatome and covered one-third or more of its surface. Nonnerve involvement was defined as a marking that did not follow the principal distribution of a dermatome or it did not cover at least one-third of the dermatome surface. The threshold values were selected by the authors to define markings that could reasonably be considered signs of nerve root involvement, based on known symptom distributions associated with nerve root involvement, that is, pain within part of a dermatome but not necessarily within the entire dermatome.

**Interrater reliability**

The three assessors independently screened the left and right body halves of 50 pain drawings, 25 from baseline and 25 from the 12-month follow-up (in random order and stratified by the time point), as described earlier. Each drawing was classified as either nerve or nonnerve involvement. In preparation, each assessor made a test assessment of one drawing with the dermatome template and the brief assessment guide;
then, any questions on the assessment guide were clarified. All assessors confirmed that they had understood how to use the guide and the requirements for an affirmative/negative finding of nerve root involvement.

Statistics

After each drawing was classified as nonnerve or nerve involvement, it was recorded as 0 or 1, respectively, in IBM SPSS Statistics 21 software. Criterion validity was analyzed as the agreement between the dermatome template classification and the clinical assessment regarding signs of nerve root involvement (WAD grade 2 or 3), evaluated with an unweighted kappa, because the data were nominal (yes/no answer) for both variables. The kappa value was characterized according to Cicchetti and Sparrow,32 where ≤0.40 was considered poor agreement, 0.40–0.59 fair agreement, 0.60–0.74 good agreement, and ≥0.75 excellent agreement. Because the kappa calculation can give unrepresentative values, in cases of an imbalance between the numbers of affirmative/negative answers,33 we supplemented the analysis with a calculation of the percentage agreement between the two methods showed an exact match in 51% of outcomes, and the expected agreement was 45%. The diagnostic accuracy of detecting nerve involvement, expressed as sensitivity and specificity, proved to be imbalanced. The sensitivity of the pain drawing was thus 93% (86 of the 92 cases assessed as WAD 3), and the specificity was 19% (23 of the 121 cases assessed as WAD 2) (Table 1).

Interrater reliability proved to be good. The Fleiss kappa for three assessors was 0.64 (standard error =0.06; 95% CI: 0.53 to 0.76; Table 2). The three assessors showed exact matches in outcome for 80% of the cases (40 drawings).

Results

The criterion validity test showed poor agreement between the pain drawing classification and the reference clinical WAD grade. The unweighted kappa was 0.11 (standard error =0.04; 95% CI: −0.03 to 0.20; Table 1).

Calculation of the percentage agreement between the two methods showed an exact match in 51% of outcomes, and the expected agreement was 45%. The diagnostic accuracy of detecting nerve involvement, expressed as sensitivity and specificity, proved to be imbalanced. The sensitivity of the pain drawing was thus 93% (86 of the 92 cases assessed as WAD 3), and the specificity was 19% (23 of the 121 cases assessed as WAD 2) (Table 1).

Interrater reliability proved to be good. The Fleiss kappa for three assessors was 0.64 (standard error = 0.06; 95% CI: 0.53 to 0.76; Table 2). The three assessors showed exact matches in outcome for 80% of the cases (40 drawings).

**Table 1** Overview of the assessment of nerve involvement among patients (n) based on a pain drawing and a clinical classification of WAD

<table>
<thead>
<tr>
<th>Pain drawing</th>
<th>Patient WAD grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WAD 2</td>
</tr>
<tr>
<td>Nonnerve involvement</td>
<td>23</td>
</tr>
<tr>
<td>Nerve involvement</td>
<td>98</td>
</tr>
<tr>
<td>Total</td>
<td>121</td>
</tr>
</tbody>
</table>

**Notes:** WAD 2, neck complaint and musculoskeletal signs; WAD 3, neck complaint and neurological signs.

**Abbreviation:** WAD, whiplash-associated disorders.

**Table 2** Calculation matrix for agreement between assessors on nerve involvement in WAD pain drawings

<table>
<thead>
<tr>
<th>Assessor 1, Physiotherapist 1</th>
<th>Assessor 2, Physician</th>
<th>Physiotherapist 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>Total</td>
</tr>
<tr>
<td>No</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>79</td>
</tr>
</tbody>
</table>

**Abbreviation:** WAD, whiplash-associated disorders.
Discussion
This study investigated the psychometric properties of a standardized assessment of the pain drawing for determining nerve involvement in the cervical spine, with the clinical assessment as reference. We found poor criterion validity, but good interrater reliability.

Diagnostic accuracy
In this study, the low specificity of the outcome indicated that the pain drawing assessment would show a high proportion of false-positive test results, when used in a similar manner and setting in clinical praxis. The poor agreement between the pain drawing assessments and the clinical physical examination was most likely influenced by the reference test used, the difficulty of completing the drawing, the validity of the dermatomes, and the chronic character of WAD.

First, it is questionable whether the clinical assessment was a reliable gold standard for diagnosing nerve root involvement. In the study sample used for the current analyses, 27,28 43% of patients were diagnosed with clinical signs of nerve root involvement in the clinical examination. However, patients with, for example, pain or sensory deficits or motor deficits could have had symptoms of nerve root involvement, although they would not fit the WAD grade 3 criteria. Moreover, it is possible that subtle symptoms and signs of nerve root involvement19,20 could have been missed in the clinical examinations of the WAD study. Both of these potential weaknesses in the clinical assessment of nerve root involvement would have caused biased results (imperfect gold standard bias) in this study, which may have affected diagnostic accuracy. However, the clinical assessment was performed by skilled physiotherapists in a structured way according to the WAD grading30 and clinical recommendations.27,30

Second, information on the neuroanatomical distribution of pain could have been omitted or inaccurately added to the drawing. For instance, the task of filling out a pain drawing could be carried out by marking a single spot (as with map pin), alternatively to shadow in a more schematic way the entire limb in question. It has been questioned whether individuals given this type of assignment possess the required competence.34 For example, varying linguistic understanding or personal abilities could hinder valid reporting. Patient thoroughness might have decreased with fatigue because the pain drawing was included in an extensive questionnaire. Conversely, completing the pain drawing might have enhanced patient engagement, and consequently, the sense of participation, due to the nonregulated form of reporting.35

Third, the peripheral projection of cervical nerve involvement may have differed from classical dermatomes or from the gold standard neuroanatomical distribution of pain with cervical nerve involvement.25,36,37 Slipman et al36 examined this issue with a diagnostic, selective nerve root block applied to the cervical spine. Immediately prior to the injection of contrast agent, a nerve root was stimulated mechanically. The patient verbally described the symptoms, which were recorded on a pain drawing. A body-sector bitmap was used, which divided the figure into 793 squares. The clinician recorded the most painful area, the second most painful area, areas that were rarely painful, and areas that were never painful. Slipman et al36 concluded that pain was experienced to a relatively large extent within the classical dermatomes but that pain outside of these areas was not unusual. This finding was considered a consequence of the many connections between the nerve roots of the cervical spine, compared to those present in the thoracic and lumbar spine. Slipman et al36 also pointed out the fact that human dermatomes were originally identified in tests of light touch, not pain or paresthesia.38,39

Fourth, assessing nerve root involvement in patients with chronic pain might have been complicated by important issues regarding pain distribution. With time, the area of pain is often enlarged due to increased irritability in the central nervous system.40 Underlying mechanisms for this phenomenon include an increased sensitivity of the secondary afferent neuron in the spinal cord dorsal horn and in neighboring postsynaptic neurons,41 activation of nociceptive neurons in close proximity, or dysfunction of the descending pain-inhibiting neural tracts. All these conditions could lead to generalized pain, when a painful condition persists over time. Moreover, some patients may have had referred or peripheral pain from the musculoskeletal system, distributed in a way that seemed to correlate with a spinal nerve (e.g., thumb arthritis), which would complicate the diagnosis. It is generally acknowledged that several physiological and psychological processes may underlie whiplash-related pain and disability.42 Part of the clinical considerations should also be the possibility of residual pain with a neuroanatomical distribution and no other neurological signs that can remain with a person years after the healing of a nerve root compromise is completed.43 Therefore, the pain would, in such cases, not be a valid criteria of ongoing nerve root involvement.

Evidence is scarce for evaluating the quality of the pain drawing with regard to identifying cervical spine nerve involvement. To our knowledge, only one study has
previously examined this issue. Bertilson et al. studied the agreement between a pain drawing assessment and a clinical examination with regard to neuropathic pain. They studied 50 Swedish patients in primary care with neck–shoulder problems, with or without radiating discomfort (median duration 9.5 months, range 9 days–60 years). The clinical examination included patient history interviews, physical examinations, and radiological reports when available, with similar diagnostic criteria for nerve involvement. They found that the pain drawing alone had 90% sensitivity for identifying neuropathic pain. Specificity was not analyzed because, in the final clinical assessment, most patients (92%) were diagnosed with nerve involvement.

**Interrater reliability**

Our results showed good interrater reliability in analyzing pain drawings for nerve root involvement. This supported the reproducibility of the clinicians’ assessment of pain/discomfort drawing, an important aspect of its usefulness in clinical praxis. This result was considered especially strong in view of the complexity of the task and the fact that reliability depends on the individual and joint training of the assessors in an observational method. In earlier studies of pain drawing as an instrument, the assessment has been made mainly by means of a grid/surface area, body regions, or dermatomes. For surface area measurement, electronic measuring is considered to provide a higher precision, and some software has been developed for this purpose. Dos Reis et al. analyzed 52 pain drawings completed by patients with chronic neck pain or neck–shoulder–arm pain. Both intrarater reliability and interrater reliability (four examiners) were found to be high, with intraclass correlations 0.99 and 0.99, respectively. The authors concluded that this demonstrated clinically acceptable interrater reliability.

Our results on interrater reliability could, in part, be attributed to the dermatome template, which served as an aid in systematic inspections of both the right and the left halves of the figure. This template displayed the entire surface areas covered by the dermatomes, knowledge that might not be immediately available.

The rating guide was succinct and straightforward, which provided a good basis for reproducibility. Despite the probability that the figure layout used in this study reduced the ability of assessors to orient themselves to the anatomy, the assessors largely agreed in their assessments; this fact may be credited to the rating guide. Drawings that were assessed differently by different assessors could not be characterized by any specific traits.

**Strengths and limitations**

Noteworthy strengths of this study were foremost that the assessors conducted the screening independently, and they were blinded, that is, they did not have access to randomization or any other patient data. Also, the large number of participants and the trained assessors provided statistically reliable results.

This study also had some limitations. First, the figure in the drawing lacked some anatomical landmarks (collar bones, navel), which may have lowered the precision of the markings, due to difficulties in transferring the patient’s physical experience to a figure. Another limitation in using the drawing as a measuring instrument was that patients were instructed to mark areas of pain but not other sensations. Alternatively, a discomfort drawing may have been more appropriate for identifying affected spinal segments because it can record all symptoms of nerve root involvement. There was some overlap between the pain drawing assessment and the clinically assessed WAD grade; however, in clinical praxis, these instruments measure different aspects and complement each other. Thus, using both methods could broaden the view of the patient’s problems. The approximate layout of the dermatome template might, at first, be perceived as a weakness. It was transferred from the original study by Lee et al. to the figure used in the WAD study questionnaire. However, this transfer was not expected to present any problems, from a methodological perspective. The foremost objective was to capture the outlines of a pain pattern, an approach that is considered appropriate for the clinical assessment of neuropathic pain. For instance, the receptive field of the affected nerve may expand over time, due to normal physiological phenomena, such as central sensitization.

**Conclusion**

Our results showed that the standardized pain drawing assessment of nerve root involvement in chronic WAD was not in agreement with the clinical assessment. Further research is warranted to optimize the utilization of a pain/discomfort drawing as an instrument for identifying nerve involvement in cervical spinal injuries.

**Acknowledgments**

The authors wish to thank the study participants and everyone who contributed to the WAD study. This study was financially supported by funding from the Swedish government in cooperation with the Swedish Social Insurance Agency, through the REHSAM foundation, the Swedish Research Council, Centre for Clinical Research Sörmland at Uppsala University.
Sweden, Uppsala-Örebro Regional Research Council Sweden, the Regional Centre for Clinical Research of Östergötland County Council, and the Medical Research Council of Southeast Sweden.

Disclosure

The authors report no conflicts of interest in this work.

References

18. Rothman SM, Winkelstein BA. Chemical and mechanical nerve root insults induce differential behavioral sensitivity and glial activation that are enhanced in combination. Brain Res. 2007;1181:30–43.