PAIN MEDICINE
Volume 4 • Number 2 • 2003

A Structured Evidence-Based Review on the Meaning of Nonorganic Physical Signs: Waddell Signs

David A. Fishbain, MD, FAPA,[a–f] Brandly Cole, PsyD,[e,f] R.B. Cutler, PhD,[a,d–f] John Lewis, PhD,[a] H.L. Rosomoff, MD, DMedSc, FAAPM,[b–f] and R. Steele Rosomoff, BSN, MBA[b–f]

Departments of [a]Psychiatry, [b]Neurological Surgery, and [c]Anesthesiology, [d]University of Miami School of Medicine, Miami, Florida; [e]University of Miami Comprehensive Pain & Rehabilitation Center at [f]South Shore Hospital, Miami, Florida

ABSTRACT

Study Design. This is a structured, evidence-based review of all available studies addressing the concept of nonorganic findings: Waddell signs (WSs).

Objectives. To determine what evidence, if any, exists for the various interpretations for the presence of WSs on physical examination.

Summary of Background Data. WSs are a group of eight physical findings divided into five categories, the presence of which has been alleged at times to have the following interpretations: Malingering/secondary gain, hysteria, psychological distress, magnified presentation, abnormal illness behavior, abnormal pain behavior, and somatic amplification. At the present time, there is, therefore, significant confusion as to what these findings mean.

Methods. A computer and manual literature search produced 61 studies and case series reports relating to WSs. These references were reviewed in detail, sorted, and placed into tabular form according to the following subject areas: 1) Reliability (test-retest); 2) Reliability (inter-rater); 3) Reliability (factor analysis); 4) Validity, psychological distress; 5) Validity, correlation Minnesota Multiphasic Pain Inventory (MMPI); 6) Validity, correlation abnormal illness behavior; 7) Validity, other behaviors; 8) Validity, as a nonorganic phenomenon; 9) Validity, correlation pain drawing; 10) Validity, functional performance; 11) Validity, treatment outcome; 12) Validity, predicting surgical treatment outcome; 13) Validity, return to work outcome; 14) Validity, secondary gain correlation; and 15) Validity, pain correlation. Each study in each topic area was classified according to the type of study it represented according to the type of evidence guidelines developed by the Agency for Health Care Policy and Research (AHCPR). In addition, a list of 14 study quality criteria was used to measure the quality of each study. Each study was categorized for each criterion as positive, (criterion filled), negative (criterion not filled), or not applicable independently by two of the authors. A percent quality score was obtained for each study by counting the total number of positives obtained, dividing by 14 minus the total number of not applicables, and multiplying by 100. Only studies having a quality score of 75% or greater were used to formulate the conclusions of this review. The strength and consistency of the evidence represented by the remaining studies in each topic area (above) was then categorized according to the strength and consistency AHCPR guidelines. Conclusions of this review for each topic area are based on these results.

Results of the Data Synthesis. Of the 61 studies, four had quality scores below 75% and were not used to generate the results of this review. According to the AHCPR guidelines for strength and consistency of the reviewed data, the following results were obtained: 1) There was consistent evidence for WSs being associated with decreased functional performance, poor nonsurgical treatment outcome, and greater levels of pain; 2) There was *generally* consistent evidence for WSs *not* being associated with psychological distress, abnormal illness behavior, or secondary gain; 3) There was also generally consistent evidence that WSs are an organic phenomenon and that they cannot be used to discriminate organic from nonorganic problems; 4) There was inconsistent evidence that

Reprint requests to: David A. Fishbain, MD, FAPA, 600 Alton Road, Miami Beach, FL 33139. Tel: (305) 532-7246; Fax: (305) 534-3974; E-mail: d.fishbain@miami.edu.

© American Academy of Pain Medicine 1526-2375/03/$15.00/141 141–181

WSs do demonstrate inter-rater reliability, *do not* correlate with the neurotic triad of the MMPI, are associated with poorer surgical treatment outcome, and are associated with nonreturn to work; 5) There was little or no evidence that WSs demonstrate test-retest reliability, or reliable factors, and are associated with self-esteem problems, catastrophizing, or the nonorganic pain drawing.

Conclusions. Based on the above results, the following conclusions were made: 1) WSs do not correlate with psychological distress; 2) WSs do not discriminate organic from nonorganic problems; 3) WSs may represent an organic phenomenon; 4) WSs are associated with poorer treatment outcome; 5) WSs are associated with greater pain levels; 6) WSs are not associated with secondary gain; and 7) As a group, WS studies demonstrate some methodological problems.

Key Words. Nonorganic Physical Signs; Waddell Signs; Nondermatomal Sensory Abnormalities; Abnormal Illness Behavior; Disease Magnification; Somatic Amplification; Conversion Symptoms; Malingering; Chronic Pain

Introduction

During the 1980s, Waddell et al. [1–3] studied a group of low back pain (LBP) physical signs frequently found in chronic pain patients (CPPs). These signs, earlier in this century, were identified as predominantly nonorganic and, as such, were thought to represent either malingering or hysteria [4]. Waddell et al. [1] standardized these signs in CPPs and placed them into five general categories (tenderness, simulation, distraction, regional, and overreaction). In all, they listed eight signs under these five general categories (Table 1). Waddell et al. [1] also determined that these signs were reliable and valid and suggested that if a CPP demonstrated three or more *categories* of nonorganic signs that the CPP should be evaluated for psychological problems.

Table 1 Waddell's nonorganic signs

Tenderness
- Superficial skin tender to light touch
- Nonanatomic deep tenderness not localized to one area

Simulation
- Axial loading pressure on the skull of a standing patient induces lower back pain
- Rotation: Shoulders and pelvis rotated in the same plane induces pain

Distraction
- Difference in straight leg raising in supine and sitting positions

Regional
- Weakness: Many muscle groups, "give-away weakness" (patient does not give full effort on minor muscle testing)
- Sensory: Sensory loss in a stocking or glove distribution, nondermatomal

Overreaction
- Disproportionate facial or verbal expression (i.e., pain behavior)

The interpretation of these nonorganic signs, henceforth called Waddell's signs (WSs), and their exact meaning has been fraught with difficulty. The presence of WSs has been thought to indicate psychological distress [3,5] and a more "magnified or more emphatic presentation of the severity of their problem" [3]. It was also proposed that WSs should draw attention to the possibility of abnormal illness behaviors [3], which was defined by Waddell [6] as "maladaptive overt illness related behavior which is out of proportion to the underlying physical disease and more readily attributable to associated cognitive and affective disturbance." WSs have also been equated with pain behavior and have been presented as a pain behavior assessment tool [6]. Finally, some other researchers have also equated WSs with "somatic amplification" [7]. Unfortunately, some of the above words, such as "magnified" and/or "out of proportion," have led some authors and clinicians [8,9] to revert to the original interpretation of WSs: Malingering or pain of psychological origin. For example, in a recent review [10] of detecting sincerity of effort, WSs are listed as one of the "widely" used methods to determine sincerity of effort. This led Main and Waddell [11] to attempt to "reappraise" the interpretation of WSs.

Since the early 1980s, WSs have been extensively used by clinicians [11]. As such, there is a wide body of literature on the subject. Unfortunately, although there has been one review [12] in this area, this literature has not been reviewed in detail in an evidence-based, structured fashion in order to clarify the interpretation of WSs. The

Table 2 Levels of evidence guidelines as applied by the AHCPR [13]

A Types of evidence guidelines
- I A meta-analysis of multiple, well-designed, controlled studies
- II At least one well-designed experimental study
- III Well-designed quasiexperimental studies, such as nonrandomized controlled, single-group pre/post, cohorts, time series, or matched case-controlled studies
- IV Well-designed nonexperimental studies, e.g., comparative, correlational, descriptive, case control
- V Case reports and clinical examples

B Strength and consistency of evidence guidelines
- A There is evidence of Type I or consistent findings from multiple studies of Type II, III, or IV
- B There is evidence of Type II, III, or IV and findings are generally consistent
- C There is evidence of Type II, III, or IV but findings are inconsistent
- D There is little or no evidence or there is Type V evidence only

Panel consensus only: Practice recommended on the basis of opinion of experts in pain management.

purpose of this evidence-based, structured review is: 1) To isolate any studies that have addressed WSs in pain patients; 2) To determine what evidence exists for the various interpretations on the meaning of WSs; 3) To evaluate the quality of that evidence; and 4) Using high-quality studies only, to evaluate the strength of that evidence through an evidence-based, structured review process using the Agency for Health Care Policy and Research (AHCPR) levels of evidence guidelines (Table 2) [13]. It is to be noted that this review should not be confused with a meta-analysis.

Methods

Relevant references were located by the following procedure. MEDLINE, Psychological Abstracts, Science Citation Index, and the National Library of Medicine Physician Data Query (PDQ) databases were searched using the following 13 subject headings: Nonorganic physical signs; Waddell signs; nondermatomal sensory abnormalities; sensory loss stocking; sensory loss glove; physical findings, distraction; nonanatomic tenderness; illness behavior; conversion symptoms; disease magnification; abnormal illness behavior; pain behavior; and somatic amplification. Each of these was exploded with the medical subject heading (MESH) "pain." Each term was exploded for subheadings in MESH, and all retrieved references were reviewed. The searches were not restricted to the English language and were conducted back to 1966, except for Science Citation Index, which was conducted back to 1974. The latest date searched for each database was 2000. A manual search was also performed using key pain journals, pain meeting abstracts, and textbooks. For the following journals, the following years were reviewed: Pain, 1975–2000; Spine, 1976–2000; Journal of Pain and Symptom Management, 1986–2000; The Pain Clinic, 1986–2000; and Clinical Journal of Pain, 1985–2000. Abstract books of the following meetings were reviewed for the following years: International Association for the Study of Pain, 1981, 1984, 1987, 1990, 1993, 1996, and 1999; and American Pain Society Meetings, 1982–2000. Three pain textbooks were reviewed for possible references. These were: Evaluation and Treatment of Chronic Pain, Third Edition, Aronoff G, editor, 1999; Handbook of Pain Management, Second Edition, Tollison CD, Satterthwaite JR, Tollison JW, editors, 1994; and Textbook of Pain, Third Edition, Wall P, Melzak R, editors, 1993. Ninety-eight references were found in this manner and were subject to a cursory review. Studies were chosen for detailed review according to one inclusion criterion only: If they addressed WSs in some fashion. Of the 98 references found, 37 either did not relate to the subject area (WSs) or used other concepts, such as nonorganic symptoms or related to disability issues. The 61 remaining studies [1,3,5,7,14–71] were sorted into groups according to which subject matter that study addressed in reference to the meaning of WSs. It is to be noted that some studies addressed more than one subject matter relevant to this review. These groups were the following: 1) Reliability (test-retest) [1,18]; 2) Reliability (inter-rater) [1,7,18–20]; 3) Reliability (factor analysis) [22]; 4) Validity, the psychological distress concept [7,22–29]; 5) Validity, correlation with Minnesota Multiphasic Personality Inventory (MMPI) [1,21,22,30,31,55]; 6) Validity, correlation abnormal illness behavior [3,7,23,29]; 7) Validity, correlation other behavior [25,32,33]; 8) Validity, as a nonorganic phenomenon [25,35–44]; 9) Validity, correlation with pain drawing [46]; 10) Validity, functional performance [27,32,33,45,47–49]; 11) Validity, treatment outcome [5,22,30,31,50]; 12) Validity, predicting surgical outcome [1,2,16,30,31,51]; 13) Validity, outcome return to work [1,14,16,30,31,52–59,68,71]; 14) Validity, secondary gain correlation [1,7,14,15,17,22,25,27,28,46,48,49,60–62]; 15) Validity, pain correlation [15,16,25,27,28,50,60,63–67,69,70].

Research information from the 61 references was then abstracted into tabular form. Appendices

1–15 were organized according to the above groupings. These tables are arranged to present the reference, research question, study design, sample size, nonorganic findings outcome measure, statistical analysis, results, quality score, categorization of type of evidence the study represented (Table 2A), and reviewer comments.

The quality of the studies was categorized according to the systems developed by and reported by Hoogendoorn et al. [72] and de Vet et al. [73]. These researchers developed and tested a list of criteria to be used to assess methodological quality of prospective, historical cohort, and case control studies. For details of how these criteria were developed, the reader is referred to the original studies [72,73]. Thirteen criteria were selected from their list that were appropriate to the studies utilized (Appendix 17). Hoogendoorn et al. [72] described 23 criteria, of which many of the criteria were not appropriate to the topic of this review and, as such, were not included in the criteria list used in this study. In addition to the 13 selected criteria, one other criterion was added. This criterion was appropriate to the topic being studied. Waddell et al. [1], in their original study, determined that, for a patient to be categorized as being positive for WSs, he/she had to have a positive finding on three out of five WS groups. As such, this was added to the criteria list as a desirable criterion. This resulted in a total of 14 criteria. The full criteria list is presented in Appendices 17A–D.

For each included study, each criterion was rated as either present/fulfilled (+), not present/ unfulfilled (–), or not applicable (NA). NA was used as follows. There were basically four types of studies analyzed for quality: Case control, cohort, correlational, and case series. Thus, some criteria in Table 2 pertained only to case control studies, while others only to cohort studies, etc. As such, NA was used if the criterion in question pertained to another type of study other than the one being reviewed. In addition, NA was used when that criterion did not pertain to the study in question, for example, no need to use WS Groups. NA was not used when information was not available or not described [74]. Under those circumstances, a negative was assigned [74]. A negative was also assigned if the item did not meet the preselected criteria [74]. Each study was rated independently for each criterion by the senior author (DF) and another author (BC). Both raters chose either a positive, negative, or NA categorization for each criterion for each study selected for detailed review. The assigned categorizations by DF and BC for each selected study were then compared in a meeting. Any discrepancies in the categorizations were resolved by mutual agreement. This resulted in a final decision as to whether each criterion received a negative, positive, or NA categorization. Categorizations were then summarized and placed into tabular format, (Appendices 17A–D). A quality score was obtained by counting he number of positives obtained. This score was divided by 14 (the total number of criteria) minus the number of NAs and multiplied by 100, which gave the percentage quality score.

Studies scoring less than 50% historically have been rated as "low quality" [74]. These studies are usually not used to arrive at conclusions about a review topic. For the purposes of this review, however, we arbitrarily set the acceptable quality score at 75%. Studies scoring less than 75% were not used in arriving at a conclusion about the reviewed topic. These studies are marked with a Q in Appendices 1–15.

The senior author was the one who independently abstracted the data into Appendices 1–15. However, data abstraction was checked independently by BC. Any discrepancies were resolved by mutual agreement. In addition, BC checked the classifications of the reviewed studies, that is, whether the reviewed study was a cohort, case control, etc. Any discrepancies in this classification were also resolved by mutual agreement.

The categorization of the type of evidence the study represented was based on the guidelines developed by the AHCPR for categorizing the levels of evidence represented by reviewed studies (Table 2A) [13]. Studies were categorized I through V according to this scheme. In this categorization, I represents a meta-analysis of well-designed, controlled studies and V represents a case report or clinical example. This categorization was also independently arrived at by the senior author and BC. Any discrepancies were again resolved by mutual agreement in a meeting format.

The strength and consistency of the research evidence in each group of studies in *each table* (Appendices 1–15) were then categorized according to the AHCPR guidelines [13] developed for this purpose (Table 2B). These guidelines allow the researcher to categorize the reviewed evidence as being consistent, generally consistent, nonconsistent, or demonstrating little evidence for supporting the hypothesis under study. In using these categorization guidelines, only studies attaining a score of 75% or greater were used. Categorizations according to these guidelines (Table 2B) were performed independently by the senior author and B.C. Any discrepancies were later resolved by mutual agreement.

In addition, all references were reviewed for the following demographic questions: 1) What percentages of chronic LBP patients demonstrated each of the individual WSs? 2) What percentages of chronic LBP patients demonstrated 0, 1, 2, 3, 4, or 5 categories of WS? 3) What percentages of CPPs with WSs demonstrated each category of WS? and 4) Is there a differential distribution of WSs between the sexes? For questions 1, 2, and 3, mean percentages were calculated if possible. These data are presented in Appendices 18–20. Finally, data from Appendices 1–15 were formatted into a summary table (Table 3). This last table was designed to summarize the overall findings of the structured review by listing the following: The research question under study in reference to WSs; categorization of the type of evidence plus the overall quality score of the evidence utilized; categorization of the strength and consistency of the evidence; and conclusions from the strength and consistency categorization.

Results

Of the 61 studies, three [18,59,60] had quality scores below 75% (Appendices 1–15). These three studies were not used to arrive at overall quality scores or to arrive at strength and consistency of evidence conclusions.

Two studies [1,18] addressed the issue of WS test-retest reliability (Appendix 1). Both studies in this group represented Type III evidence. However, one study [18] had a quality score below 75% and was therefore not used to draw conclusions. The other study [1] demonstrated test-retest reliability in 85% of patients. Based on these observations and according to AHCPR guidelines (Table 2B), the consistency of this evidence was rated D (there is little or no evidence).

Five studies [1,7,18–20] addressed the issue of WS inter-rater reliability (Appendix 2). One study [1] in this group represented Type III evidence, while the rest [7,18–20] represented Type IV evidence. One study had a quality score below 75% and was therefore not used. Two studies [1,19] supported inter-rater reliability for WSs, while two [7,20] did not. Based on this observation and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated C (inconsistent evidence).

One study [22] addressed the issue of reliability of WSs in reference to factor analysis (Appendix 3). It found that, of the eight WSs, only four could be grouped to form one reliable factor. This study represented Type III evidence. Because of a lack of other studies in this category, the consistency of this evidence was rated D (there is little or no evidence).

Nine studies [7,22–29] addressed the issue of WS correlation with psychological distress and whether this concept is valid (Appendix 4). Three studies [2,25,26] in this group represented Type III evidence, and the rest represented Type IV evidence. The majority of these studies (87.8%) supported the no-psychological-distress correlation. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated B (generally consistent).

Six studies [1,21,22,30–31,55] (one study [31] was used twice for a total of seven reports) addressed the issue of WS correlation with the neurotic triad of the MMPI and whether this concept is valid (Appendix 5). All studies in this group represented Type III evidence. Three [1,27,55] of the seven studies, or 42.8%, supported the notion that there is a correlation between WSs and the neurotic triad of the MMPI. The majority (57.2%) of the studies ([21,30,31], one study was used twice), however, supported the alternate hypothesis: No correlation. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated C (inconsistent).

Four studies [3,7,23,29] addressed the issue of WS correlation with abnormal illness behavior (AIB) and whether this concept is valid (Appendix 6). Three [7,23,29] studies in this group represented Type IV evidence and one [3] represented Type III. The hypothesis that WSs are related to AIB was not supported. All studies (100%) supported the alternate hypothesis: No correlation. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated B (generally consistent).

Three studies [25,32,33] addressed the issue of WS correlation with self-esteem [32,33] and catastrophizing [25] (Appendix 7). All studies in this group represented Type III evidence. Neither of the self-esteem studies supported the hypothesis that WSs are related to self-esteem. The one catastrophizing study indicated a relationship between WSs and catastrophizing. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated as

Table 3 Strength and consistency of the evidence for the WS research questions addressed in Appendices 3–17

Research questions	Type of evidence categorization (according to number of studies in each category) and overall quality score	Categorization of strength and consistency of the overall evidence	Conclusions from research review categorization
1. Do WSs demonstrate test-retest reliability?	1 study, Type III QS = 100%	D	Not enough studies to draw conclusion
2. Do WSs demonstrate inter-rater reliability?	1 study, Type III 3 studies, Type IV QS = 90%	C	Inconsistent evidence that WSs demonstrate inter-rater reliability
3. Do WSs demonstrate reliable factors?	1 study, Type III QS = 76.9%	D	Not enough studies to draw conclusion
4. Are WSs correlated with psychological distress?	3 studies, Type III 6 studies Type IV QS = 91.3%	B	Generally consistent findings that WSs are *not* associated with psychological distress
5. Are WSs correlated with the neurotic triad of the the MMPI?	6 studies, Type III QS = 93%	C	Inconsistent findings that WSs are *not* associated with the MMPI neurotic triad
6. Are WSs associated with abnormal illness behavior?	1 study, Type III 3 studies, Type IV QS = 100%	B	Generally consistent findings that WSs are *not* associated with abnormal illness behavior
7. Are WSs associated with: A) Self-esteem or B) Catastrophizing?	3 studies, Type III QS = 83.6%	D (self-esteem) D (catastrophizing)	Not enough studies to draw conclusions for catastrophizing and self-esteem association with WSs
8. Are WSs a nonorganic phenomenon and can WSs discriminate organic from nonorganic pain problems?	For discrimination: 2 studies, Type III 1 study, Type IV QS = 91.8% For organic phenomenon: 1 study, Type III 1 study, Type IV 4 studies, Type V	B (organic phenomenon) B (discrimination)	Generally consistent findings that WSs represent an organic phenomenon and cannot discriminate organic from nonorganic problems
9. Is there a correlation between WSs and a nonorganic pain drawing?	1 study, Type III QS = 92.3%	D	Not enough studies to draw conclusion
10. Are WSs associated with decreased functional performance?	1 study, Type IV 6 studies, Type III QS = 85.7%	A	Consistent findings that WSs are associated with poorer physical performance
11. Are WSs associated with poorer treatment outcome?	4 studies, Type III 1 study, Type IV QS = 90.5%	A	Consistent findings that WSs are associated with poorer treatment (nonsurgical) outcome
12. Are WSs associated with poorer surgical outcome?	4 studies, Type III 1 study, Type IV QS = 93.3%	C	Inconsistent findings that WSs are associated with poorer treatment (surgical) outcome
13. Are WSs associated with nonreturn to work?	14 studies, Type III 1 study, Type IV QS = 91.4%	C	Inconsistent findings that WSs are associated with poorer treatment (nonreturn to work) outcome
14. Are WSs associated with secondary gain?	13 studies, Type III 3 studies, Type IV QS = 87.5%	B	Generally consistent findings that WSs are *not* associated with secondary gains
15. Are WSs associated with pain?	5 studies, Type III 1 study, Type V 8 studies, Type IV QS = 89%	A	Consistent findings that WSs are associated with greater pain

Abbreviations: QS = average quality score of studies in this research area.

D for both self-esteem and catastrophizing (there is little or no evidence).

Ten studies [35–44] addressed the issue of whether WSs represent a nonorganic phenomenon and/or can be counted on to discriminate nonorganic from organic problems (Appendix 8). Four studies [25,35,39/40,42] addressed the issue of discrimination. However, of these, one [42] had a quality score of less than 75% and was not used. Of the remaining studies, two [25,35] were Type III and one [39/40] was Type IV. All three of those studies were consistent in indicating that WSs do not discriminate organic from nonorganic problems. Based on these observations and using

AHCPR guidelines (Table 2B), the consistency of the above evidence was rated B. Six reports [36–38,41,43,44] addressed the issue of whether WSs represent an organic rather than a nonorganic phenomenon. Four studies [36,37,43,44] in that group represented Type V evidence, while one [38] represented Type III and another [41] represented Type IV. All those studies were consistent in indicating that WSs *do* have an organic substrate. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated B (generally consistent) for both issues.

One study [46] addressed the issue of whether WSs correlate with a nonorganic pain drawing (Appendix 9). This study represented Type III evidence indicating that there was a correlation. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated D (there is little or no evidence).

Seven studies [27,32,33,45,47–49] addressed the issue of whether WSs are associated with decreased functional performance (Appendix 10). All these studies, except one [27], represented Type III evidence, while the lone study represented Type IV evidence. All these studies were consistent in indicating that WSs are associated with poorer physical performance. Based on these observations and using AHCPR guidelines (Appendix Table 2B), the consistency of this evidence was rated A (consistent).

Five studies [5,22,30,31,50] (two studies [30,31] were used twice for a total of seven reports) addressed the issue of whether WSs are associated with treatment outcome (Appendix 11). All studies, except one [5], represented Type III evidence. All of these studies, except one [22] (which was unequivocal), were consistent in indicating that WSs are associated with poorer treatment outcome. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated A (consistent).

Five studies [1,16,30,31,51] (one study was used twice for a total of six reports) addressed the issue of whether WSs are associated with poorer surgical outcome (Appendix 12). All these studies, except one [16], represented Type III evidence, while the lone study represented Type IV evidence. The studies were not consistent in indicating whether WSs are associated with poorer surgical outcome. Fifty percent of the studies [1,30,51] indicated that WSs are associated with worse surgical outcome. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated C (inconsistent).

Fifteen studies [1,14,16,30,31,52–59,68,71] (three [30,31,56] were used twice for a total of 18 reports) addressed the issue of whether WSs are associated with nonreturn to work (Appendix 13). All these studies, except one [16], represented Type III evidence, while the lone study [16] represented Type IV evidence. The studies were not consistent in indicating whether WSs are associated with nonreturn to work. Twelve studies [1,16,30,31,53–56,59,68,71], or 66.6%, indicated that WSs are associated with nonreturn to work. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated C (inconsistent).

Sixteen studies [1,7,14–17,22,25,27,28,46,48, 49,60–62] (four [1,7,25,60] were used twice and two [22,48] were used three times for a total of 24 reports) addressed the issue of whether WSs are associated with secondary gain (Appendix 14). All these studies represented Type III evidence, except for three [7,27,28], which represented type IV evidence. The majority of the studies [1,7,14,22,25, 27,28,46,48,61], 75.0%, indicated that WSs are not associated with secondary gain. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated B (generally consistent).

Fourteen studies [15,16,25,27,28,50,60,63–67, 69,70] (one [66] was used twice for a total of 15 reports) addressed the issue of whether WSs are associated with greater pain perception (Appendix 15). Five [15,25,50,60,63] studies represented Type III evidence, one study [70] represented Type V evidence, and seven studies [16,27,28,64, 66,67,69] represented Type IV. These studies were extremely consistent. All reports indicated a relationship between pain and WSs, and twelve of the reports indicated a relationship between greater pain and WSs. Based on these observations and using AHCPR guidelines (Table 2B), the consistency of this evidence was rated A (consistent).

Appendices 17A–D present the raw study quality ratings for each study for the 14 quality criteria. In addition, these tables present the overall quality score for each study.

Appendices 18–20 present the compilation of the demographic material for WSs. In reference to WS sex distribution, three studies [22,25,63] reported a higher frequency in women, while three studies [1,2,46] reported no differences between men and women.

Table 3 summarizes the strength and consistency of the evidence for the WS research questions presented in the introduction. This table indicates the following. There is substantial consistent evidence (rating A) for the following research findings: 1) WSs are associated with decreased functional performance; 2) WSs are associated with poorer nonsurgical treatment outcome; and 3) WSs are associated with greater pain levels. There is evidence that is generally consistent (rating B) for the following research findings: 1) WSs do not correlate with psychological distress; 2) WSs are not associated with abnormal illness behavior; 3) WSs do not discriminate between organic and nonorganic problems and appear to be an organic phenomenon; and 4) WSs are not associated with secondary gain. There is inconsistent evidence (rating C) for the following research findings: 1) WSs *do* demonstrate inter-rater reliability; 2) WSs *do not* correlate with the neurotic triad of the MMPI; 3) WSs are associated with poorer surgical treatment outcome; and 4) WSs are associated with nonreturn to work. There is little or no evidence (rating D) for the following research findings: 1) WSs demonstrate test-retest reliability; 2) WSs demonstrate reliable factors; 3) WSs are associated with self-esteem, catastrophizing, and a nonorganic pain drawing. In reference to the above consistent and inconsistent findings, it is to be noted that the direction of some of the inconsistent findings supports the above consistent findings. For example, the lack of correlation with the neurotic triad of MMPI supports the consistent findings of no correlation with psychological distress, abnormal illness behavior, or self-esteem. Similarly, the inconsistent findings of poorer surgical outcome and nonreturn to work support the consistent findings of poorer physical performance and poorer treatment outcome.

Discussion

Waddell et al. [1] reported WSs to demonstrate high test-retest reliability and to also demonstrate high inter-rater reliability. Unfortunately, we were only able to find one study that supported test-retest reliability, which was of poor quality. We, therefore, concluded that not enough studies exist to conclude that WSs demonstrate test-retest reliability. In addition, there was also inconsistent evidence for WS inter-rater reliability. Researchers using WSs in future studies will need to report their own inter-rater reliability figures. Similarly, clinicians using WSs to make clinical diagnoses will need to keep in mind that their findings in reference to these signs could be unreliable.

Waddell et al. [1] indicated that the "final proof of the validity of W.S. depends on their value in predicting outcome of treatment." As indicated in Table 3, the evidence is consistent (rating A) that WSs are associated with decreased functional performance and poorer nonsurgical treatment outcome. In addition, the evidence, although inconsistent (rating C), indicates that WSs are associated with poorer surgical outcome and nonreturn to work. Overall, therefore, this evidence points to the validity of WSs in predicting outcome. At issue is why WSs are associated with overall treatment outcome. This issue is discussed below in reference to the results of this evidence-based review.

The presence of WSs has been thought to indicate a magnified (nonorganic problem) [3,8,9] or to indicate the presence of secondary gain issues [8–10]. These interpretations for WSs were made in spite of the fact that Waddell et al. [3] indicated that WSs were "most related to physical severity" and that "illness behavior was secondary to physical disorder and improved or deteriorated with successful or failed treatment" [15]. The results of this evidence-based review indicate that the evidence is generally consistent (rating B) in indicating that WSs *cannot* discriminate between organic and nonorganic problems and may be an organic phenomenon. In addition, additional evidence, which is generally consistent (rating B), indicates that WSs are not associated with secondary gain issues. Thus, these two interpretations of WSs are open to question and, therefore, may not serve as an explanation for the association of WSs with treatment outcome.

Waddell's signs have been thought to be a psychological phenomenon and, in many papers, have been called behavior signs. Waddell's signs have been thought to indicate psychological distress [3,5] or abnormal illness behavior [3], to be similar to pain behavior [6], or to be somatic amplification [7]. The reviewed evidence indicates that there is very little evidence for any of these interpretations and that all these interpretations could be wrong. There is generally consistent evidence (rating B) that these signs do not indicate psychological distress or abnormal illness behavior. Additional evidence, although inconsistent (rating C), indicates that WSs are not associated with the neurotic triad of the MMPI. Thus, to date, no firm association between psychological disturbance and WSs has been established. Therefore, psychological problems may not be the reason why WSs are associated with poorer treatment outcome.

It is interesting to note that the association of pain with WSs has not been a central topic of discussion in the WS literature. This association appears to have been entirely ignored in favor of a search for psychological explanations. Nevertheless, the authors were able to locate 14 studies that directly or indirectly addressed the issue of the association of pain with WSs. This evidence was amazingly consistent (rating A) in indicating an association between greater pain and the presence of WSs. What, then, is the relationship between WSs and pain levels, and can pain actually explain the presence of WSs? It appears that there is substantial literature evidence that indicates that there may be a strong relationship between some WSs and the presence of pain. This literature is reviewed below in reference to the WSs involved.

Nondermatomal sensory loss in a stocking, glove, whole leg, or whole arm distribution is one of the regional WSs [1]. Similarly, nondermatomal or nonanatomic deep tenderness not localized to one area is another WS [1]. The central concept here is the belief that both pain/tenderness and sensory loss should occur in a dermatomal distribution [75,76]. This is a belief that has been a central tenet of neurology for many years [75]. Recently, this belief has been challenged through animal experiments and human observations. Rats with a painful peripheral neuropathy have been shown to have pain in a stocking-like distribution (nondermatomal). This has been called extraterritorial pain [77], which is related to input from undamaged nerves. Similarly, tactile hypesthesia (numbness) is often associated with painful conditions. Nathan [78] and others [79] have found that, when there is a lesion in the peripheral or central nervous system causing pain or local tenderness, stimuli applied to the skin of the affected part are not clearly perceived, that is, there is numbness. This was exactly the finding of one of the studies presented in this review [67]. Nathan [78] and others [79] found that the area of this pain-related hypesthesia could be reduced or removed if the pain was relieved. Moriwaki and Yuge [79] also reported that pain-associated tactile hypesthesia often spreads outside of the territory of the affected peripheral nerves responsible for the pain, that is, is extraterritorial. Finally, it is well known that trigger points in myofascial pain syndrome (MFPS) have been reported to refer not only pain/tenderness, but also numbness [80] and do not follow dermatomal referral patterns [81]. Indeed, nondermatomal sensory loss and nondermatomal deep tenderness are frequently found in MFPS [82,99]. Therefore, nondermatomal sensory abnormalities are consistent with referred MFPS phenomena. These observations on MFPS referral patterns recently were bolstered by one human experimental study. In that study, it was found that ongoing muscle pain could cause sensory changes in the referred pain area [83]. This short review indicates that both pain and numbness can be nondermatomal, that is, extraterritorial. In addition, MFPS can lead to nondermatomal phenomena. As such, these studies call into question the validity of a nonorganic interpretation for the WSs of nondermatomal sensory loss and nondermatomal deep tenderness. It is clear that these signs may not be a nonorganic phenomenon, as was postulated in the early literature, and may be explainable by the presence of pain.

Another WS that may not have a nonorganic etiology, but may be related to pain, is that of regional weakness. In this sign, muscle groups, also in a nondermatomal distribution, appear weak, for example, the whole arm. However, recent human evidence [84,85] has shown that acute stimulation of nociceptive muscle afferents can inhibit single motor unit activity without changing recruitment order and, thus, LBP patients may experience fatigue faster than controls [86]. In a reverse of that experiment, CPPs with apparent weakness of dorsal foot flexors were restored to normal flexor strength by pain relief [34,87]. Finally, it appears that patients with pain also experience muscle fatigue faster than comparable controls [88]. In addition, the WS of regional weakness is frequently found in patients with MFPS [82]. Those studies also point to the possibility that the WS of regional weakness may have little validity as a nonorganic sign. This sign may also be related to perceived pain and pain-associated phenomena.

Two other WSs that may not have a nonorganic etiology, but may be related to pain, are tenderness (superficial skin tenderness to light touch) and overreaction (disproportionate facial or verbal expression to examination). Recent studies demonstrated that fibromyalgia syndrome (FMS) patients, compared with controls, have greater sensitivity to pressure pain [87] and greater sensitivity to pressure pain and light touch at the site of maximal pain compared with the contralateral side [87]. Similarly, MFPS patients have been demonstrated to have qualitatively altered nociceptive processes in highly tender muscles [88]. In addition, clinicians have noted that WSs are typical features seen in FMS [89]. Thus, touch hypersensitivity (superficial skin tenderness to light touch as a WS) may be a normal clinical finding in FMS and/or MFPS. These touch hypersensitivity findings in these syndromes may indicate neuronal changes associated with central sensitization [90]. How do these observations relate to the WS of overreaction? A recent study [91] demonstrated, by multiple regression analyses, that high pain and high pain behavior (overreaction) were both independently related to tender point scores in FMS. Thus, tender point scores may be related to the superficial skin tenderness to light touch WS and to the overreaction WS. These two WSs may, therefore, have little validity as nonorganic signs and may, thus, be related to perceived pain and pain-associated phenomena.

The above discussion indicates that five WSs (nondermatomal sensory loss, nondermatomal deep tenderness, regional weakness, superficial skin tenderness to light touch, and overreaction) out of eight may not have nonorganic etiologies and may be related to pain. Thus, this evidence indirectly supports the finding in this study on the importance of the association of pain with WSs.

One of the purposes of a structured, evidence-based review is to note some systematic methodological problems in the reviewed studies. From the above discussion, it is obvious that, in clinical studies comparing patients for WSs, one needs to control for pain levels. None of the above studies demonstrated this kind of analysis. In addition, one also needs to control for the presence of FMS and MFPS or eliminate patients with these diagnoses from the analyses groups. None of the reviewed studies used this approach. There is also some evidence in the recent literature that complex regional pain syndrome (CRPS) can be related to nondermatomal sensory abnormalities [92,93]. None of the reviewed studies controlled for the presence of CRPS.

Another methodological problem noted in the reviewed studies was that of the issue of gender. The majority of the reviewed studies did not control for gender, and some studies did not report a sex break down [22]. Some studies, however, did control for gender by reporting on men only [32], while some other studies actually compared men with women for WSs [1,15, 22,25,46,63]. Some of those studies [22,25,63] reported a higher frequency of WSs in women. This finding is not surprising. There is a higher frequency of FMS within females. Thus, these findings support the possibility of a systematic error in the reviewed studies related to gender. Future WS studies will need to control for gender.

Another methodological problem noted in the reviewed studies was that of acute pain versus chronic pain. The reviewed studies varied according to the type of pain syndrome studied. For example, some studied acute LBP, some studied chronic LBP, some studied CPPs, not indicating what types of pain syndromes were involved, and finally, some studies used both acute and chronic LBP (Appendices 18–20). The issue of not controlling for the type of pain is problematic. This is because, as indicated in Appendices 18 and 19, there may be differences in the frequencies of WSs among these types of pains. As such, future WS studies will need to avoid mixing chronic and acute pain and specify clearly the groups studied. One last issue relates to WSs and chronic versus acute pain. Some authors [94] have claimed that WSs are *positive* physical findings confirming the diagnosis of chronic pain. As can be seen from Appendices 18 and 19, WSs are present in acute pain as well. As such, WSs are not positive physical findings confirming chronic pain and should not be used in this manner.

Another systematic methodological problem noted in the reviewed studies was that of nonconsistent identification of the patients exhibiting WSs. In the original study [1], Waddell et al. suggested that the presence of three out of five WS categories indicated psychological distress. Yet subsequent researchers [2,16,21,22,23,26,27,29, 30,32,33,45–50,52,54–58,60,62–66], in addition to using WS categories, have used other approaches. These have been the average numbers of WSs [14,59], a modified Waddell Score [17], some WSs only [5,18–20,24,25,30,35–44,51,53,67–70], the somatic amplification rating scale (SARS) [7], different cutoffs for WS score categories [3,22,28, 31,32], and, simply, the presence of WSs [61]. As such, this body of literature suffers from lack of consistency in using a WS method for identifying the "distressed" patient.

There are a number of problems with this review, which relate to the class of review that this review represents. There are four classes of research reviews [95]. In the first class, new developments in a field are identified and discussed. The second class of review uses empirical evidence to highlight, illustrate, or assess a particular theory or, tentatively, to propose a new theoretical framework. In the third class of review, a reviewer can organize knowledge from divergent lines of research. The fourth class of review is called the integrative type. In that type of review, the reviewer is "primarily interested in inferring generalizations about substantive issues from a set of studies directly bearing on those issues" [96]. Meta-analysis belongs to the fourth class of review. This review belongs to the third class of review [97]. As such, it is subject to the usual problems associated with traditional reviews. The problem with traditional reviews usually relates to the fact that these reviews use a vote-counting approach for summarizing outcomes of individual studies. The results of statistical tests of significance in the reviewed studies are counted, and the tally is used as a basis to draw inferences about the consistency of outcomes. This vote-counting procedure is flawed in two major ways. First, the vote-counting procedure divides the decision of the results of the studies into two groups while not providing any indication of the strength of relationships or magnitude of differences that distinguishes groups and does not necessarily take into account the quality of the reviewed studies [98]. We have attempted to control for this last problem. Second, the vote-counting method can lead to erroneous conclusions when the statistical power in individual empirical studies is low [98]. Because this review used the vote-counting method, the limits of the conclusions drawn from it should be appreciated. Definitive answers to the questions asked in this review, therefore, await the possible application of meta-analysis to the studies brought together in this review.

Conclusions

This structured, evidence-based review of the WS literature utilizing AHCPR guidelines for levels of evidence has lead to the following conclusions: 1) WSs do not correlate with psychological distress; 2) WSs do not discriminate organic from nonorganic problems; 3) WSs may represent an organic phenomenon; 4) WSs are associated with poorer treatment outcome; 5) WSs are associated with greater pain levels; 6) WSs are not associated with secondary gain; and 7) WS studies, as a group, demonstrate some methodological problems.

References

1 Waddell G, McCulloch JA, Kummel E, et al. Nonorganic physical signs in low-back pain. Spine 1980;5:117–25.
2 Waddell G. Biopsychosocial analysis of low back pain. Baillieres Clin Rheumatol 1992;6:523–57.
3 Waddell G, Pilowsky I, Bond MR. Clinical assessment and interpretation of abnormal illness behaviour in low back pain. Pain 1989;39:41–53.
4 Frost HM. Diagnosing musculoskeletal disability of psychogenic origin in orthopaedic practice. Clin Orthop 1972;82:108–22.
5 Waddell G, Bircher M, Finlayson D, et al. Symptoms and signs: Physical disease or illness behaviour? Br Med J (Clin Res Ed) 1984;289:739–41.
6 Waddell G. 1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain. Spine 1987;12632–44.
7 Korbon GA, DeGood DE, Schroeder ME, et al. The development of a Somatic Amplification Rating Scale for low-back pain. Spine 1987;12:787–91.
8 Humphreys SC, Eck JC. Clinical evaluation and treatment options for herniated lumbar disc. Am Fam Physician 1999;59:578–82.
9 Kiester PD, Duke AD. Is it malingering, or is it "real"? Eight signs that point to nonorganic back pain. Postgrad Med 1999;106:77–80, 83–4.
10 Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: A summary of methods and approaches. Phys Ther 1998;78:867–88.
11 Main CJ, Waddell G. Behavioral responses to examination. A reappraisal of the interpretation of "nonorganic signs". Spine 1998;23:2367–71.
12 Scalzitti DA. Screening for psychological factors in patients with low back problems: Waddell's nonorganic signs. Phys Ther 1997;77:306–12.
13 Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice. Clinical practice guidelines: Directions for a new program. Washington, DC: National Academy Press; 1990.
14 Polatin PB, Cox B, Gatchel RJ, et al. A prospective study of Waddell signs in patients with chronic low back pain. When they may not be predictive. Spine 1997;22:1618–21.
15 Waddell G, Morris EW, Di Paola MP, et al. A concept of illness tested as an improved basis for surgical decisions in low-back disorders. Spine 1986;11:712–9.
16 Main CJ, Wood PL, Hollis S, et al. The Distress and Risk Assessment Method: A simple patient classification to identify distress and evaluate the risk of poor outcome. Spine 1992;17:42–52.
17 Friedrich M, Gittler G, Halberstadt Y, et al. Combined exercise and motivation program: Effect on the compliance and level of disability of patients with chronic low back pain: A randomized controlled trial. Arch Phys Med Rehabil 1998;79:475–87.
18 Spratt KF, Lehmann TR, Weinstein JN, et al. A new approach to the low-back physical examination. Behavioral assessment of mechanical signs. Spine 1990;15:96–102.

19 Sobel JB, Sollenberger P, Robinson R, et al. Cervical nonorganic signs: A new clinical tool to assess abnormal illness behavior in neck pain patients: A pilot study. Arch Phys Med Rehabil 2000;81:170–5.
20 McCombe PF, Fairbank JCT, Cockersole BC, et al. Volvo Award in clinical sciences. Reproducibility of physical signs in low-back pain. Spine 1989;14:908–18.
21 Maruta T, Goldman S, Chan CW, et al. Waddell's nonorganic signs and Minnesota Multiphasic Personality Inventory profiles in patients with chronic low back pain. Spine 1997;22:72–5.
22 Lehmann TR, Russell DW, Spratt KF. The impact of patients with nonorganic physical findings on a controlled trial of transcutaneous electrical nerve stimulation and electroacupuncture. Spine 1983;8: 625–34.
23 Vendrig AA, de Mey HR, Derksen JJ, et al. Assessment of chronic back pain patient characteristics using factor analysis of the MMPI-2: Which dimensions are actually assessed? Pain 1998;76:179–88.
24 Fishbain DA, Goldberg M, Rosomoff RS, et al. Chronic pain patients and the nonorganic physical sign of nondermatomal sensory abnormalities (NDSA). Psychosomatics 1991;32:294–303.
25 Reesor KA, Craig KD. Medically incongruent chronic back pain: Physical limitations, suffering, and ineffective coping. Pain 1988;32:35–45.
26 Greenough CG. Recovery from low back pain. 1–5 year follow-up of 287 injury-related cases. Acta Orthop Scand Suppl 1993;254:1–34.
27 Novy DM, Collins HS, Nelson DV, et al. Waddell signs: Distributional properties and correlates. Arch Phys Med Rehabil 1998;79:820–2.
28 Greenough CG, Fraser RD. Comparison of eight psychometric instruments in unselected patients with back pain. Spine 1991;16:1068–74.
29 Main CJ, Waddell G. Psychometric construction and validity of the Pilowsky Illness Behaviour Questionnaire in British patients with chronic low back pain. Pain 1987;28:13–25.
30 Doxey NC, Dzioba RB, Mitson GL, et al. Predictors of outcome in back surgery candidates. J Clin Psychol 1988;44:611–22.
31 Dzioba RB, Doxey NC. A prospective investigation into the orthopaedic and psychologic predictors of outcome of first lumbar surgery following industrial injury. Spine 1984;9:614–23.
32 Hirsch G, Beach G, Cooke C, et al. Relationship between performance on lumbar dynamometry and Waddell score in a population with low-back pain. Spine 1991;16:1039–43.
33 Cooke C, Dusik LA, Menard MR, et al. Relationship of performance on the ERGOS work simulator to illness behavior in a workers' compensation population with low back versus limb injury. J Occup Med 1994;36:757–62.
34 Knutsson E, Skoglund CR, Natchev E. Changes in voluntary muscle strength, somatosensory transmission and skin temperature concomitant with pain relief during autotraction in patients with lumbar and sacral root lesions. Pain 1988;33:173–9.
35 Rolak LA. Psychogenic sensory loss. J Nerv Ment Dis1988;176:686–7.
36 Hernández-Peón R, Chávez-Ibarra, Agilar-Figueroa E. Somatic evoked potentials in one case of hysterical anaesthesia. Electroencephalogr Clin Neurophysiol 1963;15:889–92.
37 Levy R, Behrman J. Cortical evoked responses in hysterical hemianaesthesia. Electroencephalogr Clin Neurophysiol 1970;29:400–2.
38 Levy R, Mushin J. The somatosensory evoked response in patients with hysterical anaesthesia. J Psychosom Res 1973;17:81–4.
39 Fishbain DA, Goldberg M. The misdiagnosis of conversion disorder in a psychiatric emergency service. Gen Hosp Psychiatry 1991;13:177–81.
40 Fishbain DA. Response to "The misdiagnosis of conversion disorder in a psychiatric emergency service". Gen Hosp Psychiatry 1992;14:146–8.
41 Mailis A, Downar J, Nicholson K, et al. MRI in unexplainable widespread somatosensory deficits (WSDs) in patients with chronic pain. American Pain Society meeting, Atlanta, GA; 2000:158.
42 Gould R, Miller BL, Goldberg MA, et al. The validity of hysterical signs and symptoms. J Nerv Ment Dis 1986;174:593–7.
43 Rappaport M. Brain evoked potentials as a tool in psychiatric assessment. J Clin Psychiatry 1982;43: 465–7.
44 Lorenz J, Kunze K, Bromm B. Differentiation of conversive sensory loss and malingering by P300 in a modified oddball task. Neuroreport 1998;9:187–91.
45 Waddell G, Somerville D, Henderson I, et al. Objective clinical evaluation of physical impairment in chronic low back pain. Spine 1992;17:617–28.
46 Chan CW, Goldman S, Ilstrup DM, et al. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. Spine 1993;18:1717–22.
47 Menard MR, Cooke C, Locke SR, et al. Pattern of performance in workers with low back pain during a comprehensive motor performance evaluation. Spine 1994;19:1359–66.
48 Cooke C, Menard MR, Beach GN, et al. Serial lumbar dynamometry in low back pain. Spine 1992;17:653–62.
49 Waddell G, Main CJ, Morris EW, et al. Chronic low-back pain, psychologic distress, and illness behavior. Spine 1984;9:209–13.
50 Klenerman L, Slade PD, Stanley IM, et al. The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. Spine 1995;20:478–84.
51 McCulloch JA. Chemonucleolysis. J Bone Joint Surg Br 1977;59:45–52.
52 Lacroix JM, Powell J, Lloyd GJ, et al. Low-back pain. Factors of value in predicting outcome. Spine 1990;15:495–9.

53 Lancourt J, Kettelhut M. Predicting return to work for lower back pain patients receiving worker's compensation. Spine 1992;17629–40.
54 Gaines WG Jr, Hegmann KT. Effectiveness of Waddell's nonorganic signs in predicting a delayed return to regular work in patients experiencing acute occupational low back pain. Spine 1999;24: 396–400; discussion 401.
55 McIntosh G, Frank J, Hogg-Johnson S, et al. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. Spine 2000;25:147–57.
56 Kummel BM. Nonorganic signs of significance in low back pain. Spine 1996;21:1077–81.
57 Bradish CF, Lloyd GJ, Aldam CH, et al. Do nonorganic signs help to predict the return to activity of patients with low-back pain? Spine 1988;13:557–60.
58 Fritz JM, Wainner RS, Hicks GE. The use of nonorganic signs and symptoms as a screening tool for return-to-work in patients with acute low back pain. Spine 2000;25:1925–31.
59 Werneke MW, Harris DE, Lichter RL. Clinical effectiveness of behavioral signs for screening chronic low-back pain patients in a work-oriented physical rehabilitation program. Spine 1993;18: 2412–8.
60 Hayes B, Solyom CA, Wing PC, et al. Use of psychometric measures and nonorganic signs testing in detecting nomogenic disorders in low back pain patients. Spine 1993;18:1254–9.
61 Gracovetsky SA, Newman NM, Richards MP, et al. Evaluation of clinician and machine performance in the assessment of low back pain. Spine 1998;23: 568–75.
62 Rucker KS, Metzler HM, Kregel J. Standardization of chronic pain assessment: A multiperspective approach. Clin J Pain 1996;12:94–110.
63 Waddell G, Richardson J. Observation of overt pain behaviour by physicians during routine clinical examination of patients with low back pain. J Psychosom Res 1992;36:77–87.
64 Watson PJ, Poulter ME. The development of a functional task-oriented measure of pain behaviour in chronic low back pain patients. J Back Musculoskeletal Rehabil 1997;9:57–9.
65 Ohlund C, Lindstrom I, Areskoug B, et al. Pain behavior in industrial subacute low back pain. Part I. Reliability: Concurrent and predictive validity of pain behavior assessments. Pain 1994;58:201–9.
66 Dirks HF, Wunder J, Reynolds J, et al. A scale for predicting nonphysiological contributions to pain. Psychother Psychosom 1996;65:153–7.
67 Fishbain DA, Goldberg M, Ferretti T, et al. The nondermatomal sensory abnormality (NDSA) and pain perception. Pain 1990;5(suppl):S332.
68 Karas R, McIntosh G, Hall H, et al. The relationship between nonorganic signs and centralization of symptoms in the prediction of return to work for patients with low back pain. Phys Ther 1997;77: 354–60; discussion 361–9.
69 Kaziyama HHS, Teixeira TY Lin, Teixeira M. Fibromyalgia and hemisensitive syndromes. IASP 9th World Congress meeting, Vancouver, Canada; 1999:550.
70 Walters A. Psychogenic regional pain alias hysterical pain. Brain 1961;84:1–18.
71 Kool JP, Oesd PR, DeBe RA. Predictive tests for non-return to work in patients with chronic low back pain. Eur Spine J 2002;11:258–66.
72 Hoogendoorn WE, van Poppel MN, Bongers PM, et al. Systematic review of psychosocial factors at work and private life as risk factors for back pain. Spine 2000;25:2114–25.
73 de Vet H, de Bie R, van der Heijden G, et al. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997;83:284–9.
74 Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: A systematic review. Pain 1998;77:1–13.
75 Shealy CN. A physiological basis for "hysterical" analgesia. Headache 1971;11:102–6.
76 Mersky H. Regional pain is rarely hysterical. Arch Neurol 1988;45:915–8.
77 Tal M, Bennett GJ. Extra-territorial pain in rats with a peripheral mononeuropathy: Mechano-hyperalgesia and mechano-allodynia in the territory of an uninjured nerve. Pain 1994;57:375–82.
78 Nathan PW. Improvement in cutaneous sensibility associated with relief of pain. J Neurol Neurosurg Psychiatry 1960:23:202–6.
79 Moriwaki K, Yuge O. Topographical features of cutaneous tactile hypoesthetic and hyperesthetic abnormalities in chronic pain. Pain 1999;81:1–6.
80 Sola AE, Kuitert JH. Myofascial trigger point pain in the neck and shoulder girdle. Northwest Med 1955;54:980–4.
81 Simons DG, Travell JG. Myofascial origins of low back pain. Postgrad Med 1983;73:66–108.
82 Margoles MS. Clinical assessment and interpretation of abnormal illness behaviour in low back pain, G. Waddell, I. Pilowsky and M. Bond, Pain, 39 (1989) 41–53. Pain 1990;42:258–9.
83 Graven-Nielsen T, Arendt-Nielsen L, Svensson P, et al. Stimulus-response functions in areas with experimentally induced referred muscle pain—a psychophysical study. Brain Res 1997;744:121–8.
84 Svensson P, McMillan AS, Graven-Nielsen T, et al. Modulation of an inhibitory reflex in single motor units in human masseter by tonic painful stimulation. Pain 1999;83:441–6.
85 Sohn MK, Graven-Nielsen T, Arendt-Nielsen L, et al. Inhibition of motor unit firing during experimental muscle pain in humans. Muscle Nerve 2000;23:1219–26.
86 Taimela S, Kankaanpaa M, Airaksinen O. A submaximal back extension endurance test utilising

subjective perception of low back fatigue. Scand J Rehabil Med 1998;30:107–12.
87 Kosek E, Ekholm J, Hansson P. Sensory dysfunction in fibromyalgia patients with implications for pathogenic mechanisms. Pain 1996;68:375–83.
88 Bendtsen L, Jensen R, Olesen J. Qualitatively altered nociception in chronic myofascial pain. Pain 1996;65:259–64.
89 Ferrari R, Russell AS. Letter to the editor. Spine 1999;24:306.
90 Bendtsen L, Norregaard J, Jensen R, et al. Evidence of qualitatively altered nociception in patients with fibromyalgia. Arthritis Rheum 1997;40:98–102.
91 Nicassio PM, Weisman MH, Schuman C, et al. The role of generalized pain and pain behavior in tender point scores in fibromyalgia. J Rheumatol 2000;27: 1056–62.
92 Rommel O, Gehling M, Dertwinkel R, et al. Hemisensory impairment in patients with complex regional pain syndrome. Pain 1999;80:95–101.
93 Thimineur M, Sood P, Kravitz E, et al. Central nervous system abnormalities in complex regional pain syndrome (CRPS): Clinical and quantitative evidence of medullary dysfunction. Clin J Pain 1998;14:256–67.
94 Anderson TP. Positive physical findings confirming chronic pain syndrome. A clinical commentary. Am J Phys Med Rehabil 1991;70:157–8.
95 Cooper HM. Scientific guidelines for conducting integrative reviews. Rev Educ Res 1982;52:291–302.
96 Jackson GB. Methods for integrative review Rev. Educ Res 1980;50:438–60.
97 Bangert-Drowns RL. Review of developments in meta-analysis methods. Psychol Bull 1986;99:388–99.
98 Pollock VE. Meta-analysis in literature reviews. Biol Psychiatry 1993;34:345–7.
99 Rosomoff HL, Fishbain DA, Goldberg M, et al. Physical findings in patients with chronic intractable benign pain of the neck and/or back. Pain 1989;37:279–87.

Appendix 1 Reliability (test-retest)

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' Comments
Waddell et al. 1980 [1]	Are WSs stable over time?	CPPs examined for stability of WSs; 23 days (average)	50 CPPs chronic LBP	WS categories	%	WSs remained unchanged in 85%	100%	CC III	Test-retest reliability demonstrated
Spratt et al. 1990 [18] Q	Do WSs demonstrate test-retest reliability?	Patients examined for 4 WSs by different raters on different occasions	42 acute and CPPs	Trunk twist; head compression; skin tenderness; incongruence (flip)	Average intraclass correlation reliability	Average reliability 0.82	72.7%	CH III	Test-retest reliability demonstrated, but poor quality study

For key refer to Appendix 16.

Appendix 2 Reliability (inter-rater)

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Sobel et al. 2000 [19]	Are cervical nonorganic physical signs reliable?	MD and one other professional examined patients for nonorganic signs independently (eight tests)	26 CPPs	Neck nonorganic signs	Percentage agreement kappa coefficient	Agreement ranged from 100–68%; average agreement 84–86%; kappa coefficients from 1.00–0.08; five of eight signs kappas moderate strength or above	80%	COR IV	Inter-rater reliability demonstrated
McCombe et al. 1989 [20]	Are LBP WSs reliable?	Two MDs or MD and PT examined patients independently for presence of WSs	88 LBP patients (mixed?)	Individual WSs	Kappa coefficients; Pearson correlation coefficient	No sign reliable; superficial tenderness, abnormal regional sensory, and abnormal motor disturbance particularly unreliable	80%	COR IV	No inter-rater reliability
Waddell et al. 1980 [1]	Are WSs reliable?	Two MDs independently examined patients for WSs	50 CPPs	WS categories	Percentage	Over 80% reproducibility	100%	CC III	Inter-rater reliability demonstrated
Korbon et al. 1987 [7]	Are all WSs inter-rater reliable?	Two physicians independently examined patients for WSs	127 chronic LBP patients	Individual WSs making up SARS	Reliability	Axial loading, rotation, overreaction very poor inter-rater reliability; dropped from SARS	100%	COR IV	No inter-rater reliability for some WSs
Spratt et al. 1990 [18] Q	Do four WSs demonstrate inter-rater reliability?	Patients examined for four WSs by three independent examiners on two occasions	42 acute and CPPs	Trunk twist; hand compression skin tenderness; incongruence (flip)	Coefficient alpha internal consistency calculated	Alpha of 0.24 and 0.26 (low)	72.7%	CH IV	No inter-rater reliability

For key refer to Appendix 16.

Appendix 3 Reliability (factor analysis)

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Lehmann et al. 1983 [22]	Do WSs generate reliable factors?	Seven WS factors analyzed	54 chronic LBP patients	WS categories	Factor analysis	Only one reliable factor found with four tests (tenderness, rotation, axial loading, skin tenderness); loading greater than 0.4 on this factor	76.9%	CC III	According to factor analysis, some WSs do not form reliable factors

For key refer to Appendix 16.

Appendix 4 Validity: The psychological distressed concept

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Korbon et al. 1987 [7]	Do patients with high SARS scores demonstrate high SCL-90 scores?	SARS scores determined; patients completed SCL-90	127 chronic LBP patients	Individual WSs making up SARS	Chi-square	High SARS scores associated with more intense SCL-90 symptoms	100%	COR IV	Psychological distress correlation
Vendrig et al. 1998 [23]	Do WSs correlate with a psychological disturbance (distress) factor derived from the MMPI-2?	MMPI-2 data factor analyzed and four factors derived; correlations calculated between psychological disturbance (distress) factor and WSs for both men and women	240 chronic LBP patients	Total number of WSs	Pearson product moment correlation	Correlations not significant for both men and women	100%	COR IV	No psychological distress correlation
Fishbain et al. 1991 [24]	Is the presence of nonanatomical numbness (sensory loss in a stocking or glove distribution) explained by psychiatric diagnoses?	CPPs with nonanatomical numbness identified; discriminant analyses used to identify DSM-III diagnoses related to nonanatomical numbers	283 CPPs	Nonanatomical numbness as a WS	Stepwise discriminant analysis	Discriminant function identified only 22.1% of variance with psychiatric diagnoses identifying only 5.3% of variance	100%	COR IV	No psychological distress correlation

Reesor and Craig 1988 [25]	Are congruent CLBP patients characterized by more distress?	Incongruent CLBP patients (defined as: ≥2 WSs; 2) ≥3 inappropriate symptoms; 3) score >5 on pain drawing) compared with congruent CLBP patients for distress as measured by BDI	80 chronic LBP patients	Defined under study design	Discriminant analysis	BDI scale scores did not discriminate incongruent patients	84.6%	CC III	No psychological distress correlation
Greenough 1993 [26]	Do inappropriate signs distinguish psychological disturbance?	Eight psychological tests combined as "gold standard"; WSs determined; specificity and sensitivity calculated for identifying "disturbed" patients	290 chronic LBP patients	WS categories	Specificities and sensitivities	WSs not effective in evaluating psychological disturbance	100%	CH III	No psychological distress correlation
Novy et al. 1998 [27]	Do WSs correlate with elevated depression scores?	WSs and BDI scores determined on admission	75 chronic LBP patients	WS categories	Correlations	Significant correlation with BDI scores	80%	COR IV	Psychological distress correlation
Lehman et al. 1983 [22]	Does the invalid group (≥2 WSs) generate higher scores on the MMPI scales (depression, anxiety) than the valid group (zero WSs)?	WSs and MMPI scores determined	Chronic LBP patients 30 (valid) 10 (invalid)	WS categories, but different cut-off criteria from Waddell	Chi-square	Invalid group demonstrated *lower* scores on depression and anxiety	76.9%	CC III	No psychological distress correlation
Greenough and Fraser 1991 [28]	Do WSs demonstrate high sensitivity and specificity in identifying "distressed" patients as identified by a combination of eight tests?	WSs determined; seven other tests completed; distressed patients identified; WS sensitivities and specificities determined for identifying distressed patients	300 patients with both acute and chronic LBP	WS categories; different cut-off criteria from Waddell and different for men and women	Specificity sensitivity	WSs insensitive with lowest specificity	80%	COR IV	No psychological distress correlation
Main and Waddell 1987 [29]	Are the IBQ affective and hypochondriacal scales related to WSs?	WSs determined; IBQ scores taken; new IBQ scale developed; relationship calculated	200 chronic LBP patients	WS categories	Multiple regression	WSs not related to affective or hypochondriaca l scales	100%	COR IV	No psychological distress correlation

For key refer to Appendix 16.

Appendix 5 Validity: Correlation with MMPI first three scales—hypochondriasis, depression, hysteria (neurotic triad)

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality Score	Type of evidence	Reviewers' comments
Waddell et al. 1980 [1]	Do WSs correlate with MMPI scales I, II, and III (hypochondriasis, depression, hysteria) (neurotic triad)?	Pts with WSs compared with patients without WSs for MMPI neurotic triad	304 acute and chronic LBP patients	WS categories	Correlations	Low correlation with neurotic triad of MMPI	100%	CC III	Correlation with neurotic triad
Maruta et al. 1997 [21]	Do WSs correlate with MMPI scales?	Patients divided into high (3–5) and low (2–0) WS groups; compared on MMPI	507 chronic LBP patients	WS categories	Chi-square	Statistical significance found for scales 1, 3, and 8 for males, and only scale 8 for females	100%	CC III	No correlation with neurotic triad
Doxey et al. 1988 [30]	Do WSs correlate with MMPI neurotic triad?	WSs determined; MMPI scores determined	74 chronic LBP patients	WS categories	Correlations	Number of nonorganic signs correlated with MMPI scales I and III	92.3%	CC III	No correlation with neurotic triad
Dzioba and Doxey 1984 [31]	Do WSs predict pre- and posttreatment MMPI scale I, II, and III scores?	WS scores and MMPI scores determined at admission; MMPI scores at follow-up (1 year)	116 acute LBP patients	Used 21 nonorganic findings, of which two or greater were considered positive	Analysis of variance	Could not predict	90.9%	CH III	No correlation with neurotic triad
McIntosh et al. 2000 [55]	Do WSs correlate with scales I, II, and III of MMPI?	WSs and MMPI scores measured at admission	75 chronic LBP patients	Number of WSs	Pearson product moment correlations	Total number of WSs significantly correlated with MMPI scales I, II, and III	100%	CC III	Correlation with neurotic triad
Dzioba and Doxey 1984 [31]	Do WSs predict pre- and posttreatment MMPI (scales I, II, III) scores in postsurgical patients?	WS scores and MMPI scores determined at admission; MMPI scores at follow-up (1 year) in a postsurgical group	77 acute LBP patients	Used 21 nonorganic findings, of which 2 or greater were considered positive	Analysis of variance	Could not predict	90.9%	CH III	No correlation with neurotic triad
Lehman et al. 1983 [22]	Does the invalid group (≥2 WSs) generate a neurotic triad (scales I, II, and III) on the MMPI versus the invalid group (zero WSs)?	WSs determined; MMPI scores determined	Chronic LBP patients 30 (valid) 10 (invalid)	WS categories, but different cut-off criteria from Waddell	Chi-square	Invalid group was more likely to demonstrate neurotic triad	76.9%	CC III	Correlation with neurotic triad

For key refer to Appendix 16.

Appendix 6 Validity: Correlation abnormal illness behavior

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Waddell et al. 1989 [3]	Are WSs related to AIB as measured by IBQ?	WSs determined; IBQ completed	119 chronic LBP patients	Combined score: Waddell symptoms and signs	ANOVA	WSs most powerfully related to physical severity, pain, and disability DA scale; IBQ related to WSs	100%	CC III	No relationship with AIB (IBQ)
Korbon et al. 1987 [7]	Are SARS scores related to somatic amplification as measured by the SCL-90 somatization scale?	SARS scores determined; SCL-90 scores determined	127 chronic LBP patients	Individual WSs making up SARS	Chi-square	No significant differences on SCL-90 somatization scale between high and low SARS patients	100%	COR IV	No correlation with aspect of test (somatization scale) to measure AIB
Vendrig et al. 1998 [23]	Do WSs correlate with the somatic complaints factor derived from the MMPI-2?	MMPI-2 data factor analyzed and four factors derived; correlations calculated between somatic complaints factor and WSs for both men and women	248 chronic LBP patients	WS categories	Pearson product moment correlation	Correlations significant only for men not for women	100%	COR IV	No consistent correlation with aspect of test (somatic complaints factor) to measure AIB
Main and Waddell 1987 [29]	Is the new scale of the IBQ, the AHD scale related to WSs?	WSs determined; IBQ scores taken; new IBQ scale developed; relationship calculated	200 chronic LBP patients	WS categories	Multiple regressions	WSs not related to AHD scale	100%	COR IV	No correlation with aspects of test (AHD) to measure AIB

For key refer to Appendix 16.

Appendix 7 Validity: Other behavior correlations

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Hirsch et al. 1991 [32]	Do patients with high WSs have low self-esteem?	WS scores determined along with self-esteem ratings	85 men with chronic LBP	WS groups	Chi-square	Relationship not statistically significant	81.8%	CH III	No correlation between WSs and self-esteem
Cooke et al. 1994 [33]	Do WSs correlate with self-esteem?	Patients divided into high (3–5) and low (2–0) WS groups and compared on Cooper Smith Self-Esteem Test?	29 LBP patients	WS categories	Chi-square	Mean scores not significantly different	84.6%	CC III	No correlation between WSs and self-esteem
Reesor and Craig 1988 [25]	Do incongruent CLBP patients have less efficacious coping with pain?	Incongruent CLBP patients (defined as: ≥2 WSs; ≥3 inappropriate symptoms; score >5 on pain drawing) Compared with nonincongruent CLBP patients for coping	80 chronic LBP patients	Defined under study design	Discriminant analysis	Catastrophizing discriminated incongruent CLBP patients	84.6%	CC III	Correlation between WSs and catastrophizing

For key refer to Appendix 16.

Appendix 8 Validity: As nonorganic phenomena

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Rolak 1988 [35]	Is hemifacial numbness an organic or psychogenic problem?	Patients with facial numbness examined for hemifacial sensory loss; organic lesions ruled out; patients with organic lesions (reason for disorder? xcompared with those with psychogenic (no cause found)	120 patients with numbness	Numbness, nonanatomical	Chi-square comparison	No significant difference in midline splitting between groups	100%	CC III	Numbness as a WS does not discriminate organic from nonorganic
Hernandez-Peon et al. 1963 [36]	What is the mechanism of glove and sleeve analgesia in an alleged hysterical analgesia patient?	Patient underwent somatic evoked potentials of affected extremity versus nonaffected	1 patient with numbness	Numbness, nonanatomical, as a WS	NA; simple comparison; patient self-control	No evoked potential in affected extremity; nonaffected extremity normal; hysterical analgesia is a function of exaggerated sensory inhibition	NA	CR V	Numbness as a WS appears to have an organic substrate
Levy and Behrman 1970 [37]	What is the mechanism of whole arm anesthesia in an alleged hysterical patient?	Patient underwent cortical evoked potentials with stimulation of ulnar nerve in affected and nonaffected arms	1 patient with numbness	Numbness, nonanatomical, as a WS	NA; simple comparison; patient self-control	Cortical evoked potentials smaller from affected extremity	NA	CR V	Numbness as a WS appears to have an organic substrate
Levy and Mushin 1973 [38]	What is the mechanism of hemianesthesia in alleged hysterical patients?	Patients underwent cortical evoked potential with stimulation nerves on skin	9 patients with numbness	Numbness, nonanatomical; 6 patients with hemianesthesia; two full lower extremity; one stocking	NA; simple comparison; patient self-control	Cortical evoked potential smaller from affected areas	88.8%	CC III	Numbness a WS appears to have an organic substrate
Fishbain and Goldberg 1991 [39] Fishbain 1992 [40]	Does the sign of extremity paresis/ paralysis demonstrate predictive validity in terms of the diagnosis of conversion disorder?	Patients with alleged diagnosis of conversion disorder extremity paresis/paralyses followed-up	3 patients with paresis	Whole leg paresis/paralysis	NA	All three patients found to have organic disease as a cause of paresis/paralysis	100%	SUR IV	Extremity paralysis as a WS does not discriminate organic from nonorganic
Mailis et al. 2000 [41]	Are widespread sensory deficits (sensory loss in a stocking or glove distribution) usually considered "hysterical" associated with somatosensory abnormalities in the cortex as determined by FMRI?	Patients with widespread sensory deficits subjected to noxious stimuli of both anesthetic and nonanesthetic limbs during FMRI	2 patients with numbness	Numbness, nonanatomical	NA; simple comparison; patient self-control	In both patients, stimulation of anesthetic limbs resulted in bilateral primary somatosensory cortex deactivation, whereas stimulation of normal limb resulted in somatosensory cortex activation	85.7%	CC IV	Numbness as a WS appears to have an organic substrate

Gould et al. 1986 [42] Q	Are two WSs (nonanatomical sensory exam and give-away weakness) present in patients with neurological problems?	Consecutive admissions with datified neurological disease examined for WSs	30 patients with numbness or weakness	1) Nonanatomical sensory abnormalities; 2) give-away weakness	NA	29/30 (96.6%) patients demonstrated one feature of nonphysiological sensory exam and 10/30 (33.3%) had give-away weakness	66.9%	SUR IV	Neither of these signs discriminates organic from nonorganic
Reesor and Craig 1988 [25]	Do incongruent CLBP patients have less organic physical impairment than nonincongruent patients?	Incongruent CLBP patients (defined as: ≥2 WSs; ≥ 3 inappropriate symptoms; score >5 in pain drawing) compared with incongruent patients for organic physical impairment	80 chronic LBP patients	Defined under study design	Discriminant analysis	Organic physical impairment discriminated incongruent patients (should be the reverse)	84.6%	CC III	WSs do not discriminate organic from nonorganic
Rappaport 1982 [43]	Can patients with loss of sensation in an extremity (NDSA) register somatosensory stimulation in that extremity?	Patient with NDSA in one extremity; somatosensory stimulation used on that extremity	1 patient with numbness	NDSA in one extremity	NA	Awareness in cortex blocked	NA	CR V	Numbness as a WS appears to have an organic substrate
Lorenz et al. 1998 [44]	Does a patient with NDSA of the hand have a P300 response?	NDSA determined; P300 response measured	1 patient with numbness	NDSA	NA	No P300 response affected extremity versus nonaffected	NA	CR V	Numbness as a WS appears to have an organic substrate

For key refer to Appendix 16.

Appendix 9 Validity: Correlation with pain drawing

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' Comments
Chan et al. 1993 [46]	Do WSs correlate with a nonorganic pain drawing?	Pain drawings and WS scores determined; patients classified as organic or nonorganic via pain drawing; percentage of patients with high WS scores determined as organic versus nonorganic	651 acute and chronic LBP patients	WS categories	Percentage	Of nonorganic patients, 27.7% had high WS scores, while only 8.3% of the organic patients had high WS scores	92.3%	CC III	WSs may correlate with a nonorganic pain drawing; may be more frequent in WS patients

For key refer to Appendix 16.

Appendix 10 Validity: Functional performance

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Cooke et al. 1994 [33]	Are there physical activity performance differences between high WS patients and low WS patients?	LBP patients with high and low WSs compared for performance on ergos	29 LBP patients	WS categories	Chi-square	For all 13 strength variables, the low WS group performed significantly better	84.6%	CC III	WSs correlate with poorer physical performance
Novy et al. 1998 [27]	Are WSs correlated with perceived functional status?	WSs and perceived functional difficulty (Quebec Back Pain Disability Scale) determined	75 chronic LBP patients	Total number of WSs	Pearson product moment correlation	Total number of WS categories correlated with perceived functional difficulty	80%	COR IV	WSs correlate with poorer perceived functional status
Menard et al. 1994 [47]	Do patients with high WS scores demonstrate underperformance on all motor tests if there is underperformance on lumbar dynamometry?	WSs, lumbar dynamometry, and other functional tests completed; high and low WS groups compared on these tests	56 chronic LBP patients	WS categories	Chi-square	Patients with high WS scores performed significantly poorer on all tests including those not involving their painful segment (LB)	76.9%	CC III	WSs correlate with poorer physical performance
Cooke et al. 1992 [48]	Do patients with high WS scores perform worse on dynamometry than patients with low WS scores and normal volunteers?	Patients with high and low WS scores compared with normal volunteers	45 chronic LBP patients	WS categories	Chi-square	Average initial performance different for controls (higher) than low or high WS groups	76.9%	CC III	WSs correlate with poorer physical performance
Hirsch et al. 1991 [32]	Do patients with high WS scores perform worse on dynamometry than low WS score patients?	Patients with high and low WS scores compared on lumbar dynamometry performance	85 chronic LBP patients	WS categories	Multiple stepwise regression	Significant difference in performance between low and high WS categories	81.8%	CC III	WSs correlate with poorer physical performance
Waddell et al. 1993 [45]	Do physical impairment tests correlate with behavior signs?	Eight physical impairment tests were found to successfully discriminate LBP patients from normals and were related to self-reported disability: Correlations between these tests and WSs calculated	60 chronic LBP patients	WS Categories	Pearson product moment correlations	Of eight tests, only posture and lumbar flexion were completely independent of WSs	100%	CC III	WSs correlate with functional performance
Waddell et al. 1994 [49]	To what extent do WSs contribute to disability?	CPPs received a full physical and psychological assessment in order to predict disability	200 mixed acute and chronic LBP patients	WS categories	Regression	Regression answered 71.2% of variance, of which 4.5% was attributed to WSs	100%	CC III	WSs correlate with poorer physical performance

For key refer to Appendix 16.

Appendix 11 Validity: Treatment outcome

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Waddell et al. 1984 [5]	Do patients with WSs receive more treatment?	All treatments coded; WSs and pain drawing used as measures of illness behavior	380 chronic LBP patients	Illness behavior made up of pain drawing, inappropriate symptoms, and WSs	Pearson correlation coefficients and multivariate analyses	Patients with lots of illness behavior received significantly more treatment	100%	COR IV	WSs associated with more treatment
Dzioba and Doxey 1984 [31]	Do WSs predict treatment outcome as determined by orthopedic opinion and patient opinion?	WSs determined at admission and follow-up (1 year); orthopedic and patient opinions about outcome determined	116 acute and chronic LBP patients	Used 21 nonorganic findings, of which two or greater were considered positive	Analysis of variance	Both outcome variables predicted by WSs	90.9%	CH III	WSs associated with treatment outcome
Dzioba and Doxey 1984 [31]	Do WSs predict treatment outcome as determined by orthopedic or patient opinion in surgical patients?	WSs determined at admission and follow-up (1 year); orthopedic and patient opinions about outcome determined postsurgery	77 acute and chronic LBP patients	Used 21 nonorganic findings, of which two or greater were considered positive	Analysis of variance	Both outcome variables predicted by WSs	90.9%	CH III	WSs associated with treatment outcome
Doxey et al. 1988 [30]	Do WSs correlate with orthopedic outcome after surgery?	WSs determined; patients had surgery and follow-up (1 year); orthopedic outcome determined	74 acute and chronic LBP patients	? WS categories	Correlations	Number of nonorganic signs correlated with orthopedic outcome	92.3%	CC III	WSs associated with treatment outcome
Lehmann et al. 1983 [22]	Does the invalid group (≥2 WSs) have poorer outcome for acupuncture and TENS treatment than the valid group (zero WSs)?	WSs determined; treated with TENS and acupuncture; outcome scores determined	Chronic LBP patients 30 valid 10 invalid	WS categories, but different cut-off criteria from Waddell	Chi-square	1) No difference between groups as measured by average pain, MD ratings, and physical measurements; 2) Invalid patients do not respond as well to acupuncture for peak pain	76.9%	CC III	Unequivocal results
Doxey et al. 1988 [30]	Do WSs correlate with orthopedic outcome after nonsurgical treatment?	WSs determined; patients had nonsurgical treatment and follow-up (1 year); orthopedic outcome determined	74 acute and chronic LBP patients	? WS categories	Correlations	Number of nonorganic signs correlated with orthopedic outcome	92.3%	CC III	WSs associated with treatment outcome
Klenerman et al. 1995 [50]	Do WSs predict chronicity outcome (combined pain and disability)?	WSs measured along with other multiple variables; patients followed up for 1 year and measurements repeated	123 acute LBP patients	WS groups	Multiple regression	WSs plus other psychosocial variables plus present pain predicted pain and disability at 12 months	90.3%	CH III	WSs associated with treatment outcome

For key refer to Appendix 16.

Appendix 12 Validity: Prediction surgical outcome

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Waddell et al. 1980 [1]	Do WSs predict surgical outcome?	Patients with WSs compared with patients without WSs for surgical outcome?	50 chronic LBP patients	WS categories	Correlations	WSs correlated with poor surgical result	100%	CC III	WSs associated with worse surgical outcome
Main et al. 1992 [16]	Is there a relationship between surgical outcome and WSs?	Patients placed into clusters based on WSs and other variables	11 Acute and chronic LBP patients received surgery	WS cluster	None	Of seven abnormal cluster patients, five (71.4%) improved and moved into normal cluster	100%	COR IV	WSs not associated with worse surgical outcome
Doxey et al. 1988 [30]	Do WSs correlate with surgical outcome as measured by return/nonreturn to work?	WSs determined; patients had surgery and follow-up (1 year); work status determined	74 acute and chronic LBP patients	? WS categories	Correlations	Return to work did not correlate with numbers of WSs	92.3%	CC III	WSs not associated with worse surgical outcome
Dzioba and Doxey 1984 [31]	Can WSs predict surgical outcome?	WSs determined presurgery	77 acute LBP patients	Used 21 nonorganic findings, of which two or greater considered positive	Analysis of variance	Did not predict surgical outcome	90.9%	CH III	WSs not associated worse surgical outcome
McCulloch 1977 [51]	Do patients with psychogenic components have poor surgical results with chymopapain?	Patients with psychogenic pain component compared with uncomplicated organic spinal pain patients for surgical success with chymopapain	109 acute LBP patients	Psychogenic component diagnosed as: Symptoms described atypical fashion; numbness entire leg; weakness entire leg; "magnification behavior"	Percentage	11% of patients with psychogenic component successful versus 57% with uncomplicated organic spinal pain	84.6%	CC III	WSs associated with worse surgical outcome
Doxey et al. 1988 [30]	Do WSs correlate with surgical outcome as measured by orthopedic outcome?	WSs determined; patients had surgery; followed up in 1 year and orthopedic outcome rated	74 acute and chronic LBP patients	? WS Categories	Correlations	Orthopedic outcome correlated with WSs	92.3%	CC III	WSs associated with worse surgical outcome

For key refer to Appendix 16.

Appendix 13 Validity: Outcome return to work

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Lacroix et al. 1990 [52]	Can WSs predict return to work?	WSs determined at admission; follow-up at 6–24 months for work status	100 chronic LBP patients	WS categories	ANOVA; product moment correlations	Number of WSs did not predict return to work	92.3%	CC III	WSs not associated with return to work
Lancourt and Kettelhut 1992 [53]	Can WSs predict return to work?	WSs determined; follow-up at 6 months for work status	134 chronic LBP patients	Superficial palpation over response	Discriminant analys's	Superficial palpation over response included in final predictive index	84.6%	CC III	Some WSs associated with return to work
Main et al. 1992 [16]	Do patient clusters as determined by WSs relate to return to work?	CPPs placed into four clusters according to WSs and other indicators of distress; followed up	98 acute and chronic LEP patients	WS categories	Risk analysis	Greater level of distress (higher clusters) was associated with greater risk of not working	100%	COR IV	WSs associated with return to work
Gaines and Hegmann 1999 [54]	Do WSs predict a longer time before return to work (unrestricted)?	Patients with one or more WSs compared with patients without WSs for time to return to work	55 acute LBP patients	Both WS categories and number of WSs	Chi-square	One or more WSs required longer time to return to work	76.9%	CC III	WSs associated with return to work
Polatin et al. 1997 [14]	Do WSs predict return to work after a functional restoration program?	WSs determined at admission and discharge from functional restoration program; presence/absence of WSs used to predict return to work posttreatment at 1 year	50 chronic LBP patients	Number of WSs	Fisher's exact test	No statistical difference for return to work between LBP patients who increased, decreased, or had no change in WSs with treatment	90.9%	CH III	WSs not associated with return to work
Waddell et al. 1980 [1]	Do WSs predict return to work?	Patients with WSs compared with patients without WSs for return to work	50 chronic LBP patients	WS categories	Correlations	Presence of WSs correlated with nonreturn to work	100%	CC III	WSs associated with return to work
McIntosh et al. 2000 [55]	Do WSs predict time for remaining on workers' compensation benefits?	Patients on workers' compensation benefits observed for 1 year after injury for when off benefits; clinical variables used as predictors	1,752 acute LBP patients	Number of WSs	Multivariate modeling	Three or more positive WSs predictive	100%	CH III	WSs associated with return to work
Kummel 1996 [56]	Do WSs predict return to work?	Patients placed into three groups—no WSs (organic), WSs (nonorganic), WS (but organic)—return to work percentage calculated	717 chronic LBP patients	WS categories and two new signs	None	Return to work: Organic 75. 6% Nonorganic 41.2% WSs but organic 36.7%	92.3%	CC III	WSs associated with return to work
Bradish et al. 1988 [57]	Do WSs (high) predict return to work?	WSs determined at admission; work status determined at follow-up	120 acute LBP patients	WS categories	Chi-square	No significant correlation between presence of WSs at initial presentation and return to work	100%	CC III	WSs not associated with return to work
Dzioba and Doxey 1984 [31]	Do WSs predict return to work?	WSs determined at admission; work status determined at follow-up (1 year)	116 acute LBP patients	Used 21 nonorganic findings, of which two or greater considered positive	Analysis of variance	WSs predicted return to work	90.9%	CH III	WSs associated with return to work

Appendix 13 *Continued*

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Doxey et al. 1988 [30]	Do WSs correlate with return to work after nonsurgical treatment?	WSs determined; nonsurgical treatment and follow-up at 1 year; work status determined	74 acute and chronic LBP patients	? WS categories	Correlations	Return to work correlated with number of WSs	92.3%	CC III	WSs associated with return to work
Kummel 1996 [56]	Do two new nonorganic signs add to the predictability of WSs for return to work?	Patients placed into three groups: +WSs; WSs plus shoulder sign; WSs plus cervical sign	717 chronic LBP patients	WS categories plus two new signs	None	Return to work lower in patients with WSs plus other signs	92.3%	CC III	WSs associated with return to work
Dzioba and Doxey 1984 [31]	Do WSs predict return to work postsurgery?	WSs determined at admission; work status determined at follow-up (1 year); postsurgery group selected for analysis	77 acute LBP patients	Used 21 nonorganic findings, of which two or greater considered positive	Analysis of variance	WSs did not predict return to work	90.9%	CH III	WSs not associated with return to work
Fritz et al. 2000 [58]	What is the sensitivityand specificity of WSs for predicting return to work after treatment?	WSs determined; return to work/nonreturn to work determined on follow-up (1 month) after treatment	69 acute LBP patients	WS categories	Specificity, sensitivity	Low sensitivity as a screening tool; WSs not an effective tool for detecting patients at risk for prolonged work loss	81.8%	CH III	WSs not associated with return to work
Doxey et al. 1988 [30]	Do WSs correlate with return to work after surgery?	Nonorganic physical signs determined; patients had surgery and follow-up (1 year); work status determined	74 acute and chronic LBP patients	? WS categories	Correlations	Return to work did not correlate with number of nonorganic physical signs	92.3%	CC III	WSs not associated with return to work
Werneke et al. 1993 [59] Q	Do patients who do not return to work have a larger number of WSs?	WSs determined; treated for 1 month; follow-up within 3 months and work status determined; patients considered failures if did not return to work within that time or to a lower work capacity	183 ? acute and chronic LBP patients	Total number of WSs, with two or more considered positive	Chi-square	Failures were more likely to have one or more WSs; failures were also less likely to reduce their WSs by end of treatment	72.7%	CH III	WSs associated with return to work
Karas et al. 1997 [68]	Does high WS score predict return to work?	WSs determined; return to work determined at 6 months after beginning treatment	126 acute and chronic LBP patients	WSs dichotomized—high versus low	Chi-square	Patients with high WS scores less likely to return to work	76.9%	CC III	WSs associated with return to work
Kool et al. 2002 [71]	Do WSs correlate with return to work?	WSs determined; return to work determined at 1 year	99 chronic LBP patients	WS categories	Correlations	WSs correlated with return to work	100%	CH III	WSs associated with return to work

For key refer to Appendix 16.

Appendix 14 Validity: Secondary gain correlation

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Waddell et al. 1980 [1]	Do WSs correlate with WCS or with MS?	WCS and MS CPPs compared with others for frequency of WSs	304 chronic LBP patients	WS categories	Correlation	WSs equally common among controls and WCS/MS patients	100%	CC III	WSs not associated with secondary gain
Waddell et al. 1980 [1]	Are patients with WSs more likely to demonstrate MMPI elevations of validity scale scores?	CPPs with and without WSs compared on MMPI validity scale scores	304 chronic LBP patients	WSs	Correlation	No correlation between MMPI validity scores and WSs	100%	CC III	WSs not associated with secondary gain
Polatin et al. 1997 [14]	Do WSs improve? with treatment, indicating no secondary gain	CPPs evaluated at admission and at completion of treatment for WSs	50 chronic LBP patients	Total number of WSs	Paired *t*-tests	In general, all categories of WSs decreased	90.9%	CH III	WSs not associated with secondary gain
Reesor and Craig 1988 [25]	Does WCS discriminate the incongruent CLBP patient?	Incongruent CLBP patients (defined as: ≥2 WSs; ≥3 inappropriate symptoms; score >5 on pain drawing) compared with congruent CLBP patients for WCS	40 chronic LBP patients	Defined under study design	Discriminant analysis	WCS does not discriminate between the two groups	84.6%	CC III	WSs not associated with secondary gain
Lehmann et al. 1983 [22]	Does the invalid group (≥2 WSs) generate invalid scores on MMPI versus the valid group (zero WSs)	WSs determined; MMPI scores determined	Chronic LBP patients 30 (valid) 10 (invalid)	WS categories	Chi-square	No significant difference in validity scales between groups	76.9%	CC III	WSs not associated with secondary gain
Chan et al. 1993 [46]	Are WSs more frequent in patients with WCS or MS?	WSs determined; WCS and MS determined	651 ? acute and chronic LBP patients	WS categories	Chi-square	WSs not more frequent in patients with WCS or MS	92.3%	CC III	WSs not associated with secondary gain
Cooke et al. 1992 [48]	Do WSs improve with treatment?	WSs determined at beginning and end of treatment program	48 chronic LBP patients	WS categories	Percentage	Of seven patients with high WS scores, 71.4% decreased score, 4.4% increased score, and 80% stayed same	76.9%	CC III	WSs not associated with secondary gain
Cooke et al. 1992 [48]	Does self-assessment improve with treatment in patients with high WS scores?	WSs and self-assessment determined at beginning and end of program	45 chronic LBP patients	WS categories	Chi-square	Improvement in self-assessment higher for patients with high WS scores	76.9%	CC III	WSs not associated with secondary gain

Appendix 14 *Continued*

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Cooke et al. 1992 [48]	Do patients with high WS scores improve their performance on dynamometry with treatment?	WSs and dynamometry performance determined at beginning and end of treatment	48 chronic LBP patients	WS categories	ANOVA	Performance in both low and high WS groups improved significantly	76.9%	CC III	WSs not associated with secondary gain
Hayes 1993 et al. [60] Q	Do WSs predict membership in patients groups: AFC versus do not AFC?	WSs measured in patient who do and do not AFC	231 chronic LBP patients	WS categories	Discriminant analysis	WSs correctly classified 90% of patients	69.2%	CC III	WSs associated with secondary gain
Hayes et al. 1993 [60] Q	Within patient groups who do and do not AFC, do WSs correlate with inconsistency scores on the LIQ?	WSs and LIQ completed; inconsistency scores determined	231 chronic LBP patients	WS categories	Pearson correlation coefficients	For the AFC group, 0.59 correlation between LIQ and WSs; for the non-AFC group, 0.17 correlation between LIQ and WSs	69.2%	CC III	WSs associated with secondary gain
Korbon et al. 1987 [7]	Do WCS patients have higher SARS scores?	SARS completed; WCS determined	127 chronic LBP patients	Individual WSs making up SARS	Chi-square	WCS patients were found to have statistically higher SARS scores	100%	COR IV	WSs associated with secondary gain
Korbon et al. 1987 [7]	Is litigation status associated with elevated SARS scores?	SARS completed; litigation status determined	120 chronic LBP patients	Individual WSs making up SARS	Chi-square	Litigation status not associated with higher SARS scores	100%	COR IV	WSs not associated with secondary gain
Reesor and Craig 1988 [25]	Does litigation status discriminate the incongruent CLBP patient?	Incongruent CLBP patients (defined as: ≥2 WSs; ≥3 inappropriate symptoms; score >5 on pain drawing) compared with congruent CLBP patients for litigation status?	40 chronic LBP patients	Defined under study design	Discriminant analysis	Litigation status did not discriminate between groups	84.6%	CC III	WSs not associated with secondary gain
Lehmann et al. 1983 [22]	Are members of the invalid group (≥2 WSs) more likely to have a lawyer than the valid group (zero WSs)?	WSs determined; lawyer status determined	Chronic LBP patients 30 (valid) 10 (invalid)	WS categories	Chi-square	Invalid group more likely to have a lawyer	76.9%	CC III	WSs associated with secondary gain

Lehmann et al. 1983 [22]	Does the invalid group (≥2 or more WS) comply with treatment more than the valid group (zero WS)	WS determined; lawyer status determined	Chronic LBP 30 (valid) 10 (invalid)	WS categories	Chi-square	Invalid group compliant (not expected)	76.9%	CC III	WSs not associated with secondary gain
Greenough and Fraser 1991 [28]	Does WCS influence the performance of WSs in identifying "disturbed" patients as measured by a combination of eight tests?	WSs determined; seven other tests completed; disturbed patients identified; influence of WCS on identification of these patients calculated	300 both acute and chronic LBP patients	WS categories; different cut-off criteria from Waddell and different for men and women	Multivariate analysis	WCS does not influence the performance of WSs in identifying disturbed patients	80%	COR IV	WSs not associated with secondary gain
Gracovetsky et al. 1998 [61]	Do WSs assist in identifying a dishonest group of LBP patients (normals coached to simulate LBP patients and acute LBP patients coached to dissimulate, i.e. normals)?	MD evaluators used physical findings and WSs and rated patients as certainly normal, probably normal, undecided, probably abnormal, and certainly abnormal: success rate for identification as honest versus dishonest calculated	41 acute LBP patients	WS presence	Success rate calculated	Success rate 92% in honest group, but only 26% in dishonest group; WSs did not help to identify dishonest group	76.9%	CC III	WSs not associated secondary gain
Rucker et al. 1996 [62]	Do WSs correlate with an index made up of FCE level of effort and MD exaggeration questions?	WSs determined; FCE level of effort determined; exaggeration perception completed by MD	55 CPPs	WS categories	Correlation	Index significantly correlated with WSs	100%	CH III	WSs associated with secondary gain
Waddell et al. 1984 [49]	Are WSs associated with medicolegal proceedings?	CPPs received a full physical and psychological assessment in order to predict disability	200 chronic LBP patients	WS categories	Regression	Regression answered 50.1% of variance for WSs, of which 9.2% related to medicolegal proceedings	100%	CC III	WSs associated with secondary gain
Novy et al. 1998 [27]	Do patients with elevated WSs have higher MMPI validity scale scores?	Patients evaluated for WSs and with MMPI	75 CPPs	Total number of WSs	Pearson product moment correlations	No correlation between MMPI-2 validity scales and elevated WS scores	80%	COR IV	WSs not associated with secondary gain
Waddell et al. 1986 [15]	Are WSs stable?	Patients evaluated for WSs pre- and postsurgery	85 acute and chronic LBP	WS categories	Chi-square	Significant improvement in some WSs postsurgery if surgery successful (↓ pain)	90.9%	CH III	If related to secondary gain, should not change

Appendix 14 *Continued*

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical analysis	Results	Quality score	Type of evidence	Reviewers' comments
Main et al. 1992 [16]	Is there a relationship between patient clusters determined by WSs and other variables and surgical treatment outcome?	CPPs placed into four clusters according to WSs and other indicators of distress; clusters compared for surgical outcome	98 CPPs	WS categories	Percentage	Of seven patients in an abnormal cluster, five (71.4%) moved to normal cluster	100%	COR III	If related to secondary gain, should not change
Friedrich et al. 1998 [17]	Does treatment change WS scores?	WSs recorded in two groups of CLBP patients (those receiving standard PT and those receiving PT plus motivational therapy) at entrance and at 4 months	93 CPPs	Modified WSs	Chi-square	Modified WSs significantly improved in both groups at 4 months	84.6%	CC III	If related to secondary gain, should not change

For key refer to Appendix 16.

Appendix 15 Validity: Pain correlation

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical Analysis	Results	Quality score	Type of evidence	Reviewers' comments
Waddell et al. 1986 [15]	Do patients with a greater number of WSs have more pain?	Patients divided into two groups (≤2 and ≥5 WS symptoms and signs) and compared for pain	74 acute and chronic LBP patients	Combined score of Waddell symptoms and WSs	Chi-square	Patients with ≥5 symptoms/signs had more pain	90.9%	CH III	WSs associated with greater pain
Waddell and Richardson 1992 [63]	Does pain behavior correlate with WSs?	Pain behavior and WSs measured	120 chronic LBP patients	WS categories	Pearson product moment correlations	Correlation between total overt pain behavior and WSs of 0.65	100%	CC III	WSs associated with greater pain
Watson and Poulter 1997 [64]	Does pain behavior? correlate with WSs	Pain behavior and WSs measured	45 chronic LBP patients	WS categories	Correlations	Pain behavior significantly correlated with WSs, but not pain	80%	COR IV	WSs not associated with greater pain
Main et al. 1992 [16]	Is there a relationship between pain and patient clusters as determined by WSs and other measures?	CPPs placed into four clusters according to WSs and other indicators of distress. Patient clusters compared for pain	910 acute and chronic LBP patients	WS categories	Chi-square	Greater level of pain with (in greater level of distress each cluster group)	100%	COR IV	WSs associated with greater pain

Novy et al. 1998 [27]	Is there a relationship between pain and WSs?	WSs and pain measured; correlations determined	75 chronic LBP patients	Total number of WSs	Pearson product moment correlations	WSs positively and significantly correlated with pain intensity ratings	80%	COR IV	WSs associated with greater pain
Ohlund et al. 1994 [65]	Do WSs correlate with pain behavior and pain?	Patients evaluated for overt painbehavior, covert pain behavior, and pain (VAS)	103 chronic LBP patients	WS categories	Pearson product moment correlations	WSs correlate with overt and covert pain behavior and pain scores	100%	CH III	WSs associated with greater pain
Dirks et al. 1996 [66]	Does pain intensity correlate with WSs?	WSs determined and PDPPI completed by patients	114 acute and chronic LBP patients	WS categories	Simple product moment correlations; ANOVA	Correlations between PDPPI and WSs significant; ANOVA indicated that the greater mean PDPPI scores, which occurred with greater WSs, was highly significant	80%	COR IV	WSs associated with greater pain
Dirks et al. 1996 [66]	Does pain intensity correlate with WSs?	WSs determined and PDPPI completed by patients	Replicate sample of 150 LBP patients	WS categories	Simple product moment correlations; ANOVA	Correlations between PDPPI and WSs significant; ANOVA indicated that the greater mean PDPPI scores, which occurred with greater WS, was highly significant	80%	COR IV	WSs associated with greater pain
Reesor and Craig 1988 [25]	Do incongruent CLBP patients have greater pain ratings?	Incongruent CLBP patients (defined as: ≥2 WSs; ≥3 inappropriate symptoms; score >5 on pain drawing) compared for pain on MPQ with nonincongruent CLBP patients	40 chronic LBP patients	Defined under study design	Discriminant analysis	Affective pain ratings on the MPQ discriminated incongruent patients	84.6%	CC III	WSs associated with greater pain
Greenough and Fraser 1991 [28]	Does pain influence the performance of WSs in identifying "distressed" patients as measured by a combination of eight tests?	WSs determined; 70 tests completed; disturbed patients identified; influence of pain on identification of these patients calculated	300 acute and chronic LBP patients	WS categories, different cut off criteria from Waddell and different for men and women	Multivariate analysis	Pain influences the performance of WSs in identifying "distressed" patients	80%	COR IV	WSs associated with greater pain
Fishbain et al. 1990 [67]	Is the location of the WS (NDSA) related to pain in the extremity of that finding?	WS (NDSA) location determined; pain location determined; relationship between pain location and NDSA location calculated	283 CPPs	WS (NDSA)	Chi-square	Relationship between limb pain location and limb NDSA location statistically significant for all extremities	80%	COR IV	WSs associated with pain

Appendix 15 *Continued*

Reference	Research question	Study design	Sample size	Nonorganic findings measure	Statistical Analysis	Results	Quality score	Type of evidence	Reviewers' comments
Kaziyama et al. 1999 [69]	Do patients with hemibody hypalgesia have pain on affected side?	Patients with hemibody hypalgesia identified; pain side determined	29 CPPs	Hemibody hypalgesia	NA	100% had more pain on hemibody hypalgesia side	100%	COR IV	WSs associated with pain
Walters 1961 [70]	Do case reports of alleged NDSA demonstrate pain in area of sensory loss?	Sensory loss mapped; pain area mapped	2 patients	Sensory abnormalities	NA	In both patients, areas coincided	100%	CR V	WSs associated with pain
Hayes et al. 1993 [60]	Does a patient group that AFC and has a high WS score have more pain than a patient group not AFC with a low WS score?	WSs and pain levels measured in these two groups	231 chronic LBP patients	WS categories	Multivariate comparison	Significant difference on all pain measures	69.2%	CC III	WSs associated with greater pain
Klenerman et al. 1995 [50]	Do WSs predict chronicity (combined pain and disability)?	WSs determined; 1-year follow-up; disability and pain measurement taken	123 acute LBP patients	WSs	Multiple regression	WSs predicted pain and disability	90.9%	CH III	WSs associated with greater pain

For key refer to Appendix 16.

Appendix 16 Key to Appendices 1–15

VAS = visual analogue scale
PDPPI = psychometric designs pain perception inventory
ANOVA = one-way analysis of variance
* = Reduced type of evidence category because of small sample size
SUR = survey study
COR = correlational study
CC = case control study
CH = cohort study
CR = case report

PT = physical therapy/therapist
MD = physician
SARS = somatic amplification rating scale
SCL-90 = symptom checklist 90
BDI = Beck Depression Inventory
IBQ = illness behavior questionnaire
FCE = functional capacity evaluation

AIB = abnormal illness behavior
AHD = affective and hypochondriacal scale
NDSA = nondermatomal sensory abnormality
WCS = workers' compensation status
MS = medicolegal status
AFC = anticipate(s) financial compensation
LIQ = life impact questionnaire
Q = because quality score << 75%, this study was not used to draw overall conclusions
DSM-III =
DA =
FMRI =
TENS =
MPQ =

Appendix 17A Quality of studies fulfilling inclusion criteria as rated for 14 methodological criteria

Criteria	Criteria for which type of study	[1] CC	[18] CH	[19] COR	[20] COR	[7] COR	[23] COR	[24] COR	[25] CC	[26] CH	[27] COR	[22] CC	[28] COR	[29] COR	[21] CC	[30] CC	[31] CH
1. Positive if the study had a clearly defined objective	CH/CC	+	+	NA	NA	NA	NA	NA	+	+	NA	+	NA	NA	+	+	+
2. Positive if the main features (description of the sampling frame, distribution of the population according to age and sex) of the study population were described	CH/CC	+	+	NA	NA	NA	NA	NA	+	+	NA	+	NA	NA	+	+	+
3. Positive if the participation rate at baseline was at least 80%	CH	NA	+	NA	NA	NA	NA	NA	NA	+	NA	NA	NA	NA	NA	NA	+
4. Positive if cases and controls were drawn from the same population and clear definitions of cases and controls were given	CC	+	NA	NA	NA	NA	NA	NA	+	NA	NA	+	NA	NA	+	+	NA
5. Positive if the participation rates of cases and controls selected and invited to participate at baseline were at least 80%	CC	+	NA	NA	NA	NA	NA	NA	+	NA	NA	–	NA	NA	+	+	NA
6. Assessment of back pain or neck pain: Positive if based on standardized methods of acceptable quality	CH/CC	+	+	NA	NA	NA	NA	NA	+	+	NA	+	NA	NA	+	+	+

Appendix 17A *Continued*

Criteria	Criteria for which type of study	[1] CC	[18] CH	[19] COR	[20] COR	[7] COR	[23] COR	[24] COR	[25] CC	[26] CH	[27] COR	[22] CC	[28] COR	[29] COR	[21] CC	[30] CC	[31] CH
7. Positive if exposure was measured in an identical manner among cases and controls	CC	+	NA	NA	NA	NA	NA	NA	+	NA	NA	+	NA	NA	+	+	NA
8. Assessment of back pain or neck pain: Positive if the exposure was measured in an identical manner among cases and controls	CH/CC	+	+	NA	NA	NA	NA	NA	+	+	NA	+	NA	NA	+	+	+
9. Analysis and data presentation: Positive if the method used for the statistical analysis was appropriate for the outcome studied	CH/CC	+	+	NA	NA	NA	NA	NA	+	+	NA	+	NA	NA	+	+	+
10. Description of inclusion and exclusion criteria		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
11. Restriction to a homogenous study population (i.e., only chronic or acute pain)		+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+
12. Allocation procedure not leading to bias		+	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+
13. Smallest group bigger than 50 subjects		+	–	–	+	+	+	+	–	+	+	–	+	+	+	+	+
14. Used 3/5 Waddell group categories as positive		+	–	+	+	+	+	–	–	+	–	–	–	+	+	–	–
Quality score		13/13 100%	8/11 72.7%	4/5 80%	4/5 80%	5/5 100%	5/5 100%	5/5 100%	11/13 84%	11/11 100%	4/5 80%	10/13 76.9%	4/5 80%	5/5 100%	13/13 100%	12/13 92.3	10/11 90.9%

Abbreviations: CC = case control study; CH = cohort study; COR = correlational study; ES = experimental study; + = present/fulfilled; – = not present/not fulfilled.

Appendix 17B Quality of studies fulfilling inclusion criteria as rated for 14 methodological criteria

Criteria	Criteria for which type of study	[55] CH	[3] CH	[32] CH	[33] CC	[35] CC	[38] CC	[39,40] Survey	[41] CC	[42] Survey	[46] CC	[47] CC	[48] CC	[49] CC	[45] CC	[5] COR
1. Positive if the study had a clearly defined objective	CH/CC	+	+	+	+	+	+	NA	+	NA	+	+	+	+	+	NA
2. Positive if the main features (description of the sampling frame, distribution distribution of the population according to age and sex) of the study population were described	CH/CC	+	+	+	+	+	+	NA	+	NA	+	+	+	+	+	NA
3. Positive if the participation rate at baseline was at least 80%	CH	+	NA	–	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4. Positive if cases and controls were drawn from the same population and clear definitions of cases and controls were given	CC	NA	+	NA	+	+	NA	NA	NA	NA	+	+	+	+	+	NA
5. Positive if the participation rates of cases and controls selected and invited to participate at baseline were at least 80%	CC	NA	+	NA	+	+	+	NA	NA	NA	+	–	–	+	+	NA
6. Positive if data were collected by means of standardized methods of acceptable quality	CH/CC	+	+	+	+	+	+	NA	+	NA	+	+	+	+	+	NA

Appendix 17B *Continued*

Criteria	Criteria for which type of study	[55] CH	[3] CH	[32] CH	[33] CC	[35] CC	[38] CC	[39,40] Survey	[41] CC	[42] Survey	[46] CC	[47] CC	[48] CC	[49] CC	[45] CC	[5] COR
7. Positive if exposure was measured in an identical manner among cases and controls	CC	NA	+	NA	+	+	+	NA	+	NA	+	+	+	+	+	NA
8. Assessment of back pain or neck pain: Positive if based on standardized methods of acceptable quality	CH/CC	+	+	–	+	+	NA	NA	NA	NA	+	–	–	+	+	NA
9. Analysis and data presentation: Positive if the method used for the statistical analysis was appropriate for the outcome studied	CH/CC	+	+	+	+	+	+	NA	NA	NA	–	+	+	+	+	NA
10. Description of inclusion and exclusion criteria		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
11. Restriction to a homogenous study population (i.e., only chronic or acute pain)		+	+	+	–	NA	NA	NA	NA	NA	+	+	+	+	+	+
12. Allocation procedure not leading to bias		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
13. Smallest group bigger than 50 subjects		+	+	+	–	+	–	+	–	–	+	–	–	+	+	+
14. Used 3/5 Waddell group categories as positive		+	+	+	+	NA	NA	NA	NA	NA	+	+	+	+	+	+
Quality score		11/11 100%	13/13 100%	9/11 81.8%	11/13 84.6%	11/11 100%	8/9 88.8%	3/3 100%	6/7 85.7%	2/3 66.6%	12/13 92.3%	10/13 76.9%	10/13 76.9%	13/13 100%	13/13 100%	5/5 100%

Abbreviations: CC = case control study; CH = cohort study; COR = correlational study; ES = experimental study; + = present/fulfilled; – = not present/not fulfilled.

Appendix 17C Quality of studies fulfilling inclusion criteria as rated for 14 methodological criteria

Criteria	Criteria for which type of study	[50] CH	[16] COR	[51] CC	[52] CC	[53] CC	[54] CC	[14] CH	[56] CC	[57] CC	[58] CH	[59] CH	[60] CC	[61] CC	[62] CH	[15] CH
1. Positive if the study had a clearly defined objective	CH/CC	+	NA	+	+	+	+	+	+	+	+	+	+	+	+	+
2. Positive if the main features (description of the sampling frame, distribution of the population according to age and sex) of the study population were described	CH/CC	+	NA	+	+	+	+	+	+	+	+	+	+	+	+	+
3. Positive if the participation rate at baseline was at least 80%	CH	–	NA	NA	NA	NA	NA	+	NA	NA	+	–	NA	NA	+	+
4. Positive if cases and controls were drawn from the same population and clear definitions of cases and controls were given	CC	NA	NA	+	+	+	+	NA	+	+	NA	NA	+	+	NA	NA
5. Positive if the participation rates of cases and controls selected and invited to participate at baseline were at least 80%	CC	NA	NA	+	+	+	+	NA	+	+	NA	NA	+	+	NA	NA
6. Positive if data were collected by means of standardized methods of acceptable quality	CH/CC	+	NA	+	+	+	+	+	+	+	+	+	+	+	+	+

Appendix 17C *Continued*

Criteria	Criteria for which type of study	[50] CH	[16] COR	[51] CC	[52] CC	[53] CC	[54] CC	[14] CH	[56] CC	[57] CC	[58] CH	[59] CH	[60] CC	[61] CC	[62] CH	[15] CH
7. Positive if exposure was measured in an identical manner among cases and controls	CC	NA	NA	+	+	+	+	NA	+	+	NA	NA	−	+	NA	NA
8. Assessment of back pain or neck pain: Positive if based on standardized methods of acceptable quality	CH/CC	+	NA	+	−	−	−	+	+	+	−	−	+	−	+	+
9. Analysis and data presentation: Positive if the method used for the statistical analysis was appropriate for the outcome studied	CH/CC	+	NA	−	+	+	+	+	−	+	+	+	+	+	+	+
10. Description of inclusion and exclusion criteria		+	+	+	+	+	+	+	+	+	+	+	−	+	+	+
11. Restriction to a homogenous study population (i.e., only chronic or acute pain)		+	+	+	+	+	+	+	+	+	+	+	−	+	+	−
12. Allocation procedure not leading to bias		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
13. Smallest group bigger than 50 subjects		+	+	+	+	+	−	+	+	+	+	+	+	−	+	+
14. Used 3/5 Waddell group categories as positive		+	+	−	+	−	−	−	+	+	−	−	−	−	+	+
Quality score		10/11 90.9%	5/5 100%	11/13 84.6%	12/13 92.3%	11/13 84.6%	10/13 76.9%	10/11 90.9%	12/13 92.3%	13/13 100%	9/11 81.8%	8/11 72.7%	9/13 69.2%	10/13 76.9%	11/11 100%	10/11 90.9%

Abbreviations: CC = case control study; CH = cohort study; COR = correlational study; ES = experimental study; + = present/fulfilled; − = not present/not fulfilled.

Appendix 17D Quality of studies fulfilling inclusion criteria as rated for 14 methodological criteria

Criteria	Criteria for which type of study	[17] CC	[63] CC	[64] COR	[65] CH	[66] COR	[67] COR	[69] COR	[70] CR	[68] CC	[71] CH
1. Positive if the study had a clearly defined objective	CH/CC	+	+	NA	+	NA	NA	NA	NA	+	+
2. Positive if the main features (description of the sampling frame, distribution of the population according to age and sex) of the study population were described	CH/CC	+	+	NA	+	NA	NA	NA	NA	+	+
3. Positive if the participation rate at baseline was at least 80%	CH	NA	NA	NA	+	NA	NA	NA	NA	NA	+
4. Positive if cases and controls were drawn from the same population and clear definitions of cases and controls were given	CC	+	+	NA	NA	NA	NA	NA	NA	+	NA
5. Positive if the participation rates of cases and controls selected and invited to participate at baseline were at least 80%	CC	+	+	NA	NA	NA	NA	NA	NA	–	NA
6. Positive if data were collected by means of standardized methods of acceptable quality	CH/CC	+	+	NA	+	NA	NA	NA	NA	+	+

Appendix 17D *Continued*

Criteria	Criteria for which type of study	[17] CC	[63] CC	[64] COR	[65] CH	[66] COR	[67] COR	[69] COR	[70] CR	[68] CC	[71] CH
7. Positive if exposure was measured in an identical manner among cases and controls	CC	+	+	NA	NA	NA	NA	NA	NA	+	NA
8. Assessment of back pain or neck pain: Positive if based on standardized methods of acceptable quality	CH/CC	+	+	NA	+	NA	NA	NA	NA	–	+
9. Analysis and data presentation: Positive if the method used for the statistical analysis was appropriate for the outcome studied	CH/CC	+	+	NA	+	NA	NA	NA	NA	+	+
10. Description of inclusion and exclusion criteria		+	+	+	+	+	+	+	NA	+	+
11. Restriction to a homogenous study population (i.e., only chronic or acute pain)		+	+	+	+	+	+	+	+	+	+
12. Allocation procedure not leading to bias		+	+	+	+	+	+	+	NA	+	+
13. Smallest group bigger than 50 subjects		–	+	–	+	+	+	+	+	–	+
14. Used 3/5 Waddell group categories as positive		–	+	+	+	–	–	NA	NA	+	+
Quality score		11/13 84.6%	13/13 100%	4/5 80%	11/11 100%	4/5 80%	4/5 80%	4/4 100%	2/2 100%	10/13 76.9%	11/11 100%

Abbreviations: CC = case control study; CH = cohort study; COR = correlational study; ES = experimental study; + = present/fulfilled; – = not present/not fulfilled.

Appendix 18 Mean percentage values (range of studies) for each WS of total population studied for chronic pain and acute pain

Waddell Sign	Mean pain percentages reported (chronic pain) [1,14,63]	Mean percentages reported (acute pain) [1,54]
Superficial tenderness to light touch	24.7%	9.3%
Nonanatomic deep tenderness	26.4%	7.8%
Axial loading	31.4%	16.0%
Rotation	21.9%	13.6%
Distraction	14.6%	1.5%
Give-away weakness	22.4%	9.1%
Sensory stocking nondermatomal	26.1%	2.4%
Overreaction	24.3%	13.8%

Appendix 19 Mean percentages of WS categories for chronic pain compared with acute pain

WS category	Mean percentages reported for chronic pain [27,32,33,46,47]	Mean percentages reported for acute pain [58]
Zero categories	39.0%	43.4%
One category	21.6%	28.9%
Two categories	15.2%	11.6%
Three categories	11.9%	5.8%
Four categories	13.9%	8.6%
Five categories	2.9%	1.4%
Three plus categories [1,21,32,46,55]	27.4%	9.4%

Appendix 20 Mean percentages of WS categories of those chronic pain patients with WSs

WS category	Mean percentages of WS categories [57]
Tenderness	32.5%
Simulation	52.7%
Distraction	50.1%
Regional	44.4%
Overreaction	46.3%