SUBJECTIVE OUTCOME ASSESSMENTS FOR CERVICAL SPINE PATHOLOGY: A NARRATIVE REVIEW

DIANE N. RESNICK, DC^a

^aAssistant Professor, Clinical Education Coordinator, University Health Center Pasadena, Southern California University of Health Sciences. Submit requests for reprints to: Dr. Diane N. Resnick, University Health Center, Pasadena, 1450 N. Lake Avenue, Pasadena, CA 91104, Office: 626-798-7805, E-mail: dianeresnick@scuhs.edu. Paper submitted February 23, 2005. Sources of Support: No funds were received for the preparation of this manuscript.

ABSTRACT

Background: Subjective outcome assessment instruments may be used to determine whether changes occur as a result of a particular treatment intervention. Since 1991, 11 outcome assessment instruments either specific to or validated for use with cervical spine pathologies have been developed and their psychometric properties critically assessed. Though a systematic review of this subject was published in 2002, it included an analysis of only 5 measures available at that time.

Objective: To present a description of each of the 11 measures and briefly compare their psychometric properties, reliability, validity, and responsiveness to change.

Methods: Computer-based searches of MEDLINE and CINAHL were performed to capture all data relevant to the eleven outcome assessment tools currently available.

Results: Data regarding descriptions, reliability and validity of 11 outcome measures were found and collated.

Conclusion: The choice of which outcome measurement to use becomes a matter of preference based on the particular patient population each clinician treats, the ease of use for the patient and ease of scoring for the clinician, and the domains that are most relevant to a particular patient. (J Chiropr Med 2005;4:113–134)

Key Indexing Terms: Neck Pain; Outcome Assessment; Questionnaires; Pain Measurement; Psychometrics; Cervical Spine

INTRODUCTION

Given the pervasiveness of work environments requiring either prolonged static postures or repetitive motions, it is not surprising that neck pain occurs frequently in industrialized societies. Lifetime prevalence rates for neck pain are high, ranging between 67% and 71%.^{1,2} In the United States, the point prevalence of neck pain was estimated at almost 50%.³ It is also estimated that in 30% of these cases, patients will develop chronic pain lasting greater than 6 months.^{4,5}

If the primary goals of treatment for cervical spine disorders are the reduction of pain and restoration of function, then outcome measures provide the lens through which clinicians may view their treatment choices and determine if these and allied goals have been attained. Historically, the real impetus to develop some measure of accountability for decisions related to patient care came about in the 1980s following attempts to control ever-expanding costs within the Medicare system.⁶ Despite radical changes, costs continued to escalate, prompting investigations into the variability of medical interventions. The primary objective of these studies was to determine which methods of treatment were most cost effective for the insurance carrier as well as efficacious for the patient.^{7–9} By the late 1980s, the health care industry witnessed a marked expansion in the development and use of outcome assessment tools, particularly those measures that focused on patients' subjective assessments of their experience.

In general, subjective outcome measures are used over time to infer causality between a particular intervention and subsequent observable changes.¹⁰ They may be broadly classified into generic, condition-specific, region-specific, and patient-specific instruments. Generic measures, such as the MOS SF-36¹¹ and the Dartmouth COOP Charts¹² assess a multitude of social, psychological and physical parameters and, due to the inclusiveness of their design, are often lengthy in nature. This translates into

a time-consuming activity for both the patient to complete and the clinician to score. However, they are particularly useful for patients with comorbidities as a more narrowly-focused instrument may overlook the effect of a secondary condition on treatment outcomes.¹³ Condition-specific and region-specific measures, such as the Oswestry Low Back Pain Disability Index¹⁴ and the Neck Disability Index¹⁵ assess factors directly related to the condition or body region, respectively. They may be used to compare the health status of patients with similar complaints or affected body regions. Compared to generic measures, condition-specific measures are considered superior with respect to content validity, responsiveness to change and in their ability to capture the health status of a patient.^{16–18} For patient-specific measures such as the Patient-Specific Functional Scale,¹⁹ each patient is asked to develop a list of activities that are negatively affected by their condition. This list of activities is unique for each patient and therefore cannot assist the clinician in comparing treatment effects among patients.

Each category of outcome measure may assess a variety of possible domains or dimensions including cognitive-behavioral aspects, disability, function, generic health status, impairment, nonorganic components, pain severity, patient satisfaction with care, physical symptoms, psychological distress, socioeconomic issues, well-being and quality of life.^{13,16,20} For clarification, impairment implies any loss or abnormality of anatomic, physiologic, mental, or emotional structure or function while functional limitation corresponds to a restriction in performance of the whole person. Disability is a term used to describe the underperformance of a socially defined role or task.²¹

Item selection, wording and scaling varies from measure to measure. For instance, 1 measure of low back pain and disability may use a list of questions that elicit either a “yes” or “no” response from the patient while another measure may ask the patient to estimate their level of pain, agreement with a particular statement or rating of a queried functional activity by circling a number from 0 to 10 or by placing a mark on a line 10 centimeters in length, similar to the visual analog scale. Yet another scale may instruct the patient to give a global response such as “improved,” “no change,” “not improved,” to a series of questions and still another scale may present 4 or 5 statements for each activity of interest and ask the patient to choose the 1 that best de-

scribes his or her current level of functional capability. For outcome measures in general and musculoskeletal assessment tools in particular, the diversity of domains and scaling options available to researchers has resulted in a lack of outcome measure standardization and an inability to directly compare treatment effects between trials.²⁰

While a thorough review of statistical concepts is beyond the scope of this paper, a brief review will assist the reader in understanding the strengths and weaknesses of the individual outcome measures described within the body of this article. Numerical values assigned to answers on a subjective outcome measure are considered variables. They may be classified as categorical, ordinal or interval, depending on the represented “value” of the variable. Categorical data represents variables with no intrinsic order to their value. For example, hair or eye color and type of shoe style would be classified as categorical data. Ordinal variables are similar to categorical variables but there is a clear ordering with respect to values assigned. For instance, a rating system of *excellent*, *good*, *fair*, *poor* would fall within the ordinal scale. While it is clear that there is a highest and lowest category, the difference between categories, with respect to actual value, is variable and potentially inconsistent; ordinal values are not equally spaced. If the values were equally spaced, the data would be considered an interval variable. In this case, there is a definite measurable value between variables. The hash marks on a ruler, thermometer, or speedometer represent interval variables. The type of variable determines which statistical computations are appropriate.

An instrument is said to be reliable if it measures something in a reproducibly consistent manner.²² The items constituting the scale should measure the same construct, thereby demonstrating internal consistency, expressed as Cronbach’s coefficient or the coefficient alpha. A low value for Cronbach’s alpha indicates that the data is multidimensional. A reliability coefficient of 0.7 to 0.8 is considered satisfactory when comparing groups while Cronbach’s alpha should be 0.9 or higher for clinical applications.²³ Intraclass correlation coefficient is also used as a measure of the reliability of a variable and is typically employed to assess inter-rater reliability. If the value is large, there is very little variation between groups. Temporal stability implies that, given a stable phenomenon, the scores generated by an

instrument for 1 patient on 2 different occasions should not differ.

If a scale measures what it is designed to measure, it is said to be valid. For the purposes of this review, we will discuss 4 primary types of validity: face, construct, concurrent, and criterion-related. A scale demonstrates face validity if the components of that scale seem to assess the chosen construct. Taking face validity a step further by adding depth, construct validity implies that the measure is comprehensive in its scope, that it measures all the significant parameters of a given construct. A scale is said to demonstrate concurrent validity if the scale, or components of the scale for multidimensional tools, correlate with the findings of similar assessment instruments.²⁴ Last, criterion-related validity implies that some secondary variable consistently and predictably correlates with the scores of a scale.

Pearson's *r*, also known as the product-moment, represents the line which best fits the scatter plot of data points. Spearman's rho is the most common correlation for use with 2 ordinal variables or an ordinal and an interval variable. For ranked data, Spearman's *r* equals Pearson's *r*. Unlike the ICC which measures the reliability of a variable, Pearson's *r* measures the validity of the variable. The value of Pearson's, intraclass, and Spearman correlation coefficients are usually similar for the same set of data.

In addition to properties of reliability and validity, it is important that a given outcome measure demonstrate responsiveness to clinical change. The more responsive an instrument is, the greater its ability to detect smaller degrees of clinically significant change. The smallest amount of change detectable between subsequent measures is known as the minimal detectable change. This parameter may not be clinically significant to either the clinician or the patient. To address this issue, Jaeschke and colleagues²⁵ proposed the idea of minimal clinically important differences wherein the change must have beneficial meaning to the patient as well as the clinician.

Both the effect statistic and the reliable change index can be used to determine the magnitude of change in units of standard deviation or mean difference divided by standard error of measure, respectively.²⁶ The standard error of measurement is

defined as the error associated with a measurement in the same units as the measure.²⁷ For example, the standard error of measure for the 101-point Numerical Rating Scale would be noted in that scale's point system. An important detail to remember is that the measurement error inherent in each outcome instrument negatively impacts the degree of responsiveness of the scale. If a 3-point difference is required between measures to confidently say a meaningful change has occurred, but the inherent measurement error is ± 3 points, then the clinician cannot determine whether a change has actually taken place; any treatment decisions based on this data may be inappropriate. The receiver operating characteristic curve can also describe the presence of change.²⁸ In this case, the sensitivity and specificity of change are plotted on a Cartesian grid. The area under the resultant curve correlates directly with the ability of the measure to distinguish which patients actually underwent meaningful change. Being able to understand the statistical and psychometric merits of 1 outcome measure versus another ensures appropriate application in the clinical setting. Bombardier¹³ developed a list (Table 1) of user-friendly questions to ask when assessing the merits of an outcome measure for use in the clinical setting.

Since 1991, a variety of subjective outcome measures applicable to the cervical spine have been

TABLE 1
KEY QUESTIONS TO ASK ABOUT AN
OUTCOME MEASURE¹³

CONTENT, POPULATION/SETTING, PURPOSE
1. WHICH OUTCOMES DO YOU WANT TO MEASURE?
2. WHICH POPULATION WILL YOU BE STUDYING AND IN WHAT SETTING?
3. WHAT IS THE PURPOSE OF YOUR STUDY – TO DESCRIBE, PREDICT, OR TO MEASURE CHANGE?
CONTENT VALIDITY
4. WHAT DOMAINS AND ITEMS ARE INCLUDED?
5. ARE THERE IMPORTANT OMISSIONS OR INAPPROPRIATE INCLUSIONS?
FACE VALIDITY
6. IS EACH QUESTION PHRASED IN A SUITABLE WAY?
7. ARE THE RESPONSE CATEGORIES APPROPRIATE?
8. IS THERE AN OVERALL SCORE SUMMARIZING ACROSS QUESTIONS? HOW IS IT CALCULATED?
FEASIBILITY
9. IS IT EASY TO UNDERSTAND?
10. IS IT EASY TO USE?
11. IS IT ACCEPTABLE?
12. WHAT FORMAT IS IT AVAILABLE IN?
13. HOW LONG DOES IT TAKE TO ADMINISTER?

TABLE REPRINTED WITH PERMISSION FROM: BOMBARDIER C. OUTCOME ASSESSMENTS IN THE EVALUATION OF TREATMENT OF SPINAL DISORDERS. SUMMARY AND GENERAL RECOMMENDATIONS. SPINE 2000;25:3100–3103.

developed^{15,16,19,29-36} and their psychometric properties critically assessed. In 2002, Pietrobon and colleagues³⁷ published a systematic review that included 5 functional outcome measures for cervical pain or dysfunction. The review included data published through the year 2000. Since then, several researchers have presented either original instruments or modifications of existing measures. The purpose of this narrative review is to identify subjective outcome measures specific to the cervical spine, including those published since the year 2000, and to offer the reader a comparison of their psychometric properties as well as a presentation of the results of formal investigations into their meaningfulness and utility.

METHODS

Computerized bibliographic databases of medical and allied health literature including MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane, and the Physiotherapy Evidence Database (PEDro) were searched in sequence. All searches were limited to human and English-language studies. There was a publication date restriction of January 1966 to April 2004 for Medline searches and 1985 to 2004 for CINAHL searches. In PEDro, articles chosen related to the cervical spine. For the other databases, the search terms included the following alone and in combination: *cervical spine, disability evaluation, functional assessment, neck pain, outcome assessment, outcome measurement, psychometrics, questionnaire, reliability, responsiveness, scale, and validity*. Additional articles were identified from references of relevant articles and assessed for inclusion in the review. A hand search was performed of all issues of the *Journal of Rehabilitation Outcomes Measures* from 1997 to the last available issue.

Each study identified through the literature search was evaluated for inclusion in the review against the following criteria: studies of surgical and non-surgical interventions for cervical spine pathology, wherein any psychometric property of an outcome measure or comparison of 2 or more outcome measures was reported; studies which investigated scale performance of outcome assessment scores as given by normal participants.

All studies consisted of adult subjects. The types of studies considered for review included meta-analyses, systematic reviews, randomized controlled trials, cohort studies and case series. Case studies,

commentaries, non-English-language articles and unpublished data were excluded from the review.

RESULTS

To date, there are 8 English-language subjective outcome assessment tools specific to the cervical spine and 3 non-specific measures that have been validated for use with neck pain patients that are available for review in the literature. Some of these measures were developed from existing instruments that assessed either similar conditions or different body regions. In addition to English-language instruments, several authors have developed versions for use in non-English speaking populations.³⁸⁻⁴¹ The earliest neck-specific measure to appear in a peer-reviewed journal was the Neck Disability Index,¹⁵ published in 1991. Since that landmark paper, new measures have consistently appeared for consideration. Table 2 lists these measures in chronological order by date of publication.

Bournemouth Neck Questionnaire

The Bournemouth Neck Questionnaire (BNQ)²⁹ (Appendix A), a minimally modified form of the Bournemouth Questionnaire,²⁰ was developed to assess the affective domains of anxiety and depression, cognitive aspects including beliefs about pain control, activities of daily living (ADL), recreational and work capabilities and pain severity in patients with non-specific neck pain. The instrument consists of 7 questions wherein the patient is asked to rate the effect of her or his neck pain using a 0 to 10 numerical rating scale. Essentially, the questions are those of the original Bournemouth Questionnaire

TABLE 2
CHRONOLOGY OF CERVICAL SPINE SUBJECTIVE
OUTCOME ASSESSMENTS

OUTCOME ASSESSMENT	PUBLICATION DATE
NECK DISABILITY INDEX	1991
NORTHWICK PARK NECK PAIN QUESTIONNAIRE	1994
DISABILITY RATING INDEX	1994
PATIENT-SPECIFIC FUNCTIONAL SCALE	1998
COPENHAGEN NECK FUNCTIONAL DISABILITY SCALE	1998
NECK PAIN AND DISABILITY VISUAL ANALOG SCALE	1999
FUNCTIONAL RATING INDEX	2001
EXTENDED ABERDEEN SPINE PAIN SCALE	2001
BOURNEMOUTH NECK QUESTIONNAIRE	2002
CERVICAL SPINE OUTCOMES QUESTIONNAIRE	2002
WHIPLASH DISABILITY QUESTIONNAIRE	2004

with 3 wording changes under the ADL scale to reflect activities affecting the cervical rather than the lumbar spine.

The validation study was conducted within 9 facilities using 102 new or existing patients presenting with a new complaint of non-specific neck pain. Patients filled out the BNQ at intake and before leaving that same day to assess test-retest reliability. Additionally, participants completed the Neck Disability Index (NDI), the Copenhagen Neck Functional Disability Scale (NFDS) and the SF-36 to provide data used to determine external construct validity. Face and content validity were not judged in this study and were assumed to be similar to that of the Bournemouth Questionnaire for the back.

Results of the study demonstrated that there was moderate test-retest reliability with intra-class coefficient (ICC) scores ranging between 0.50 and 0.63 for individual questions and 0.65 for the total score. External construct validity was tested pre- and post-treatment and overall good correlation was found between the BNQ and appropriate scales of the NDI, NFDS and SF-36. For pretreatment patients, Pearson's r was greatest between the BNQ and the NFDS ($r = 0.62$) for Item 3 assessing recreational, social and family activities. For post-treatment patients, the correlation was strongest between the BNQ and the NFDS pain intensity scale. Overall, correlation was greater between the BNQ and the NFDS pre-treatment and between the NBQ and the NDI post-treatment. Regarding responsiveness, all of the individual BNQ change scores were above the 0.3 cutoff advocated by Stratford et al,⁴² ranging between 0.42 (Item 5) and 0.82 (Item 2).

Cervical Spine Outcomes Questionnaire

The Cervical Spine Outcomes Questionnaire (CSOQ)¹⁶ consists of 33 questions assessing a variety of psychometric properties associated with complaints of neck, shoulder and arm pain. The authors identified over 100 variables used to characterize neck disorders and treatment outcomes and reviewed similar questionnaires for lumbar spine disorders. Fifteen experts evaluated this list of variables for content validity. Once consensus was reached, the questionnaire was administered to 50 patients with neck complaints to determine if any important items were overlooked.

For their validation study, the authors used a convenience sample of 216 patients under consideration for cervical spine surgery. In general, patients were Caucasian men and women, between 40 to 60 years of age, who had experienced neck pain intermittently for 6 years. Almost 60% of participants presented with cervical radiculopathy, 1 in 5 with cervical myelopathy and slightly fewer with axial neck pain. Ninety-four percent of participants demonstrated spondylitic abnormalities. Approximately two thirds of these patients had received prior conservative treatment for their neck complaints and almost 1 in 5 had undergone prior cervical surgery. While it is noted that, of these 216 patients, only those selected for anterior fusion were considered eligible for inclusion in this study, the exact number of patients completing the CSOQ was not stated.

The questionnaire is composed of 6 measures encompassing 2 health care utilization questions, 12 pain severity questions, 7 functional disability questions, 6 psychological distress questions, and 5 physical symptoms questions. The 2 pain measures use a 6-point adjective rating scale (*none, mild, discomforting, distressing, horrible, and excruciating*) developed by Melzak and Torgerson.⁴³ The functional disability and psychological distress measures use a 4-point adjective rating scale. All scores elicited from adjective rating scales were converted to numerical values, either 0–5 or 0–3 depending on the measure, and then rescaled to yield composite scores for each measure, ranging from 0 to 100. The physical symptoms and health care utilization measures consisted of yes-no questions; the number of affirmative answers were totaled for each measure and then rescaled to range between 0 and 100 points. Additionally, drug use was determined and participants were given a score of 1 each for psychoactive or narcotic drugs were used more than twice per week.

The results of the initial validation study suggest that, except for the health care utilization measure, the CSOQ is reliable and valid. Cronbach's alpha homogeneity coefficients, used to assess internal consistency, ranged between 0.80 and 0.86 for all measures except health care utilization, indicating questions within each particular measure were highly intercorrelated. Construct validity was assessed by comparing scores of the CSOQ with those of the Oswestry Disability Questionnaire (ODQ) and the SF-36. The CSOQ scores were highly correlated

with the appropriate scores of both the ODQ and the SF-36. A comparison of treatment satisfaction scores between improved and unimproved participants and effect size values indicate that the neck pain, functional disability and physical symptoms measures were highly responsive and the shoulder-arm pain and psychological distress measures were moderate responsive.

There are several points to consider with respect to this outcome measure. There are 2 questions that participants either consistently did not answer or answered incorrectly: sexual activity and number of hours in an average 24-hour period spent doing nothing. Also, the participant pool was solely comprised of surgical patients; the degree of responsiveness may not translate to non-surgical patients in the primary care setting. The authors did not discuss how each question was weighted. For instance, the 2 pain measures comprised 12 questions and were scored on a 6-point scale while the 6 psychological distress questions were scored on a 4-point scale, yet each measure was rescaled to a 0 to 100-point scale. Last, the authors did not compare their measure to any of the 8 existing assessment tools applicable to the cervical spine.

Copenhagen Neck Functional Disability Scale

The Copenhagen Neck Functional Disability Scale (CNFDS)³¹ comprises 15 questions that relate to the patient's ADL as well as psychosocial components of pain. Patients are asked to mark their responses as either "yes," "occasionally," or "no." The CNFDS was developed by a group of rheumatologists, physiotherapists and the primary investigator and was tested on 39 patients who had undergone surgery for cervical disc herniation, 21 patients seeking treatment for chronic neck pain and 117 patients receiving conservative treatment for mechanical neck pain. The instrument was also tested in a previous trial, though the appropriate reference was not given. While the authors indicate that this measure was specifically designed to overcome NDI's lack of applicability within European populations that typically do not drive, the 2 measures differ significantly in how questions are posed to patients, the types and numbers of descriptors used and the addition of psychosocial dimensions in the CNFDS.

The CNFDS displayed short-term test-retest reliability with Pearson's product moment correlation coefficients ranging between $r = 0.96$ and $r = 0.99$.

ICCs ranged between 0.95 and 0.99. Cronbach's coefficient for entire scale was reported as 0.897. Individual Cronbach's alpha scores measuring internal consistency of questions ranged between 0.88 and 0.90. The authors did find that women tended to score higher on questions 2 and 15, indicating significant bias in terms of gender for those questions. Content validity was assumed by a comparison of positive responses on the CNFDS with those of the NDI and the Northwick Park Neck Pain Questionnaire. However, the authors did not administer either of these 2 outcome measures as part of their validation process. Construct validity of the CNFDS disability scores as compared to the patient's pain scores was $r = 0.83$ ($p < 0.0001$), indicating moderate construct validity for that parameter. The authors found greater correlation between the patient's global assessment and CNFDS scores ($r = 0.89$) than with CNFDS scores and the doctor's global assessment ($r = 0.56$). Responsiveness was assessed at the end of the treatment period and at 4- and 12-month follow-up sessions by comparing changes in CNFDS scores with changes in self-reported pain scores; the results indicated strong and continuous positive correlation. However, the responsiveness of the scale may be restricted by the limited number of response possibilities.

Disability Rating Index

Developed in the early 1990s, the Disability Rating Index (DRI)³⁴ (Appendix B) was designed to assess the extent of disability associated with common physical impairments in a user-friendly format. Though primarily developed for the assessment of patients with back pain, the aim was expanded to include patient's suffering from neck and shoulder pain as well. The questionnaire consists of 12 items concerning physical function associated with ADL, work capacity and physical activity/recreation. Responses are recorded on a 100-millimeter visual analog scale with 0 = *without difficulty* and 100 = *not at all*. The clinician may use the individual measurements to create a detailed profile and the mean of all measurements as an overall index.

There were 10 cohorts within the study, 1 of which was a neck/shoulder pain group ($n=127$; 45 females, 82 males) and 5 of which consisted of neck/shoulder/low back pain patients ($n=176$; 115 females, 61 males). Patients were drawn from industrial health care centers, insurance offices, and rehabilitation units. The median age for the participants in these cohort groups was 40.5 years.

With reference to the 6 cohorts described above, test-retest reliability of the DRI was assessed only with 1 of the neck/shoulder/low back pain cohorts. An overall correlation coefficient of 0.95 was achieved while individual item coefficients ranged from 0.70 to 0.92. Inter-rater and intra-rater reproducibility scores were 0.99 and 0.98, respectively. Cronbach's alpha was 0.84, demonstrating acceptable internal consistency. Construct validity was analyzed in a variety of ways. The authors assessed whether the DRI could differentiate levels of disability, yield similar scores in different patient groups, positively correlate with hypothesized effects of increasing age on disability scores, and achieve similar outcome scores when compared to the Functional Status Questionnaire (FSQ) and the ODQ. Varying degrees of construct validity were found. Of note, the authors found that the DRI corresponded more closely with the ADL scales of the FSQ than the ODQ (ICC = 0.46 and 0.38, respectively), though the degree of correlation was low for both. Face validity was determined by evaluating the percentage of correct responses given by patients and healthy subjects (98.2% and 90.9%, respectively).

Unfortunately, responsiveness of the DRI was evaluated in only 1 cohort group of arthritis patients. Unlike the majority of patients included in the neck/shoulder and neck/shoulder/low back groups, the arthritis group was assessed pre- and post-surgery. The degree of change in this group is anticipated to be much greater than that expected in a chronic pain group. Therefore, the responsiveness of the DRI with respect to the other 9 cohorts cannot be determined.

Comparisons of DRI scores between individuals with the same disorder may prove difficult as the quantity of an activity (eg, what constitutes "*heavy work*" or how long is a "*longer sitting time*") is not specified. However, this does increase the specificity of the DRI to each patient. While the outcome measure may be used to assess patients with neck pain, the clinician must recognize that comorbid factors, particularly those affecting the low back or knees, will likely affect the individual and overall scores obtained with this instrument.

Extended Aberdeen Spine Pain Scale

In its original format, the Aberdeen Back Pain Scale (ABPS)⁴⁴ was designed to measure health status and disability of low back pain patients. Several years

after initial development, the scale was expanded and renamed the Extended Aberdeen Back Pain Scale³⁶ (EABPS) to allow application to other regions of the spine. The instrument comprises 28 questions of which 18 are relevant to the cervical spine. Half of the original ABPS questions remain on the extended version while others are modified or changed completely to accommodate adaptation to the entire spine. The expanded questionnaire was completed by 512 consecutive patients. While gender and age ranges are given, data is not provided regarding inclusion or exclusion criteria.

The authors gave the questionnaire to 10 patients to check for clarity of wording and 15 health care professionals to assess for face validity. Both groups agreed that the questions were understandable and appropriate. The authors chose to describe their results in terms of confidence interval analysis rather than correlation coefficients, citing that the later may lead to exaggerated degrees of agreement between scores. Test-retest mean differences in scores were 4.7 (SD 9.2) for the neck respondents. Cronbach's alpha, measuring internal consistency was 0.9 for neck scores. The authors suggest that the assessment instrument displays construct validity as scores improved following treatment by 1 standard deviation or greater. Pearson's correlation coefficient was higher with physical scores of the SF-12 than with mental scores. Regarding responsiveness, the measure showed large changes when health status improved and smaller changes when health status deteriorated. The complete measure may be found in the referenced article.

Functional Rating Index

Noting that patients often present with complaints in both the neck and back regions, Feise and colleagues developed the Functional Rating Index (FRI)³⁰ instrument that attempts to quantitatively measure subjective perceptions of pain and daily function in patients with spinal conditions. At that time, the only subjective outcome assessment tools applicable to cervical spine complaints were the NDI, the Northwick Park Neck Pain Questionnaire (NPNPQ) and the DRI. In addition to assessing frequency of pain, the authors selected 9 scale items from the NDI and the ODQ. Following a review of the subject, the authors determined that a 5-point Likert scale was the easiest and most useful format for patients to use and clinicians to score. However, unlike the NPNPQ or NDI, which employ a list of 5

to 6 possible answers per item, respectively, with corresponding check boxes, the FRI uses a visual analog-like scale divided into 4 regions, each with a corresponding descriptor. Possible scores range from 0 to 4, with the later indicating greater pain or dysfunction. The index score is determined by summing up the scores, dividing by the total possible (40 points) and converting to a percentage.

One hundred thirty 9 ambulatory patients from 8 chiropractic clinics who were seeking care for spinal pain and dysfunction were included in this study. Patients were assigned to 1 of 4 cohorts testing reliability, validity, responsiveness or clinical utility. Exclusion criteria included subjects with reading limitations and those requiring translation services. The mean age for subjects was 41 years, 60% of patients were women, and the vast majority (89.3%) were Caucasian. None of the demographic parameters showed any significance in terms of effect on total FRI score.

The average index score for the FRI was 47.9%. Test-retest reliability was assessed in 1 cohort (n=51) by having patients complete the questionnaire at intake and, on average, 1.95 days later (range 1–11 days). The ICC was 0.99 and Cronbach's alpha was 0.92 for all variables. Spearman's rho (0.54–0.82) indicated good item efficiency. All items were moderately correlated. A second cohort of 24 participants was used to assess the validity of the instrument. The FRI was found to correlate strongly with the SF-12 Physical Component Score ($r=0.76$) and the DRI ($r=0.76$). There was only a weak correlation between the FRI and the SF-12 Mental Component Score. The authors chose not to compare the FRI with a cervical spine-specific outcome measure such as the NDI. Content validity was determined by soliciting feedback on the domains included in the FRI from doctors and patients in 12 different chiropractic clinics. The third cohort assessing responsiveness totaled 72 patients, of which only 36 had cervical spine complaints. The standard response mean after 1 month of chiropractic treatment for cervical spine patients was 1.26, indicating a large size effect.²⁸ The fourth cohort of 28 patients was used to determine clinical utility. The FRI was correctly completed by 88% of patients.

Neck Disability Index

The NDI¹⁵ (Appendix C) was the first outcome assessment instrument developed specifically to eval-

uate the impact of neck pain on ADL. The authors modified the wording of the ODQ, replaced references to the back with those for the neck and added appropriate domains such as headache pain and frequency. The result was a 10-item questionnaire that measures neck pain levels, personal care, lifting, reading, headaches, concentration, work capacity, driving, sleeping and recreation. Using a Likert scale format, patients are instructed to mark the most appropriate of 6 possible responses. Items are scored from 0 to 5 and the total scale score is determined by totaling the individual item scores, multiplying by 2 and converting to a percentage.

The initial validation study was conducted on 48 patients (17 males, 31 females) presenting with neck pain, 34 of whom had sustained recent whiplash-type trauma. Test-retest reliability was assessed by administering the NDI to the first 17 patients at intake and 2 days later during a no-treatment period. Pearson's product moment correlation coefficient was 0.89, indicating a high degree of reliability. Internal consistency of individual items on the NDI, as determined by calculating Cronbach's alpha, ranged from 0.77 to 0.84; overall, the total index alpha coefficient score was 0.80. Construct validity was roughly assessed by plotting mean scores on a frequency histogram and comparing the resultant distribution to the clinical population. The authors suggested that the near normal bell distribution equated to evidence of construct validity. Concurrent validity was determined by comparing scores given on the NDI with those given on a visual analog scale (n=10) measuring pain as well as the McGill Pain Questionnaire (MPQ)⁴⁵ (n=30), considered the gold standard for pain measurement by the authors. Pearson's product moment correlation coefficient for pre- and post-treatment scores between the NDI and the VAS was 0.60 and between the NDI and the MPQ was 0.70, indicating moderate correlation between the NDI and both the VAS and MPQ.

The NDI was also the subject of validation studies by other authors. Hains and colleagues⁴⁶ found stable psychometric properties as evidenced by high internal consistency ($\alpha = 0.92$). In an investigation of the content and construct validity of the NDI, the NPNPQ and a problem elicitation technique for patients suffering from whiplash-associated disorders, Hoving et al⁴⁷ demonstrated that the NDI and NPNPQ were moderately correlated ($r = 0.88$, $P < 0.01$), indicating construct validity. Regarding content validity, the authors observed that the NDI did

not assess emotional function, a frequent and important issue in the population studied. In a comparison of the NDI and both mental and physical component scores of the SF-36, Riddle and Stratford¹⁸ found that the NDI demonstrated more correlation with SF-36 scores than with cervical spine active ranges of motion. Based on this data, the authors surmised that both outcome assessment tools were appropriate for use with neck pain patients.

Neck Pain and Disability Visual Analog Scale

Prior to its development, none of the existing 5 subjective outcome measures pertinent to the cervical spine adequately assessed the multidimensional effects of neck pain. The Neck Pain and Disability Visual Analog Scale (NPAD)³⁵ consists of 20 questions related to pain severity, ADL, work and social activities, difficulty with active neck motion and several questions concerning concentration, emotions, depression and effects on family relationships. Patients are instructed to place an "X" along a segmented 100 mm visual analog scale to indicate their present status. End points are indicated with 0 = not at all and 5 = worst possible situation their pain problem has taken them. It is unclear exactly how the scale is scored. The authors indicate that scores range from 0 to 5 but do not state whether the distance between the patient's mark and the end point of zero is measured or if a general score of 0 to 5 is given dependent on which of 5 sections the patient's mark is placed.

One hundred consecutive patients (61 female, 39 male; average age 44 years) who presented with neck pain were selected for the validation process. Eighty-two percent of participants reported a duration of neck pain greater than 3 months. A control group of 52 patients with low back and leg pain and 27 volunteers without neck pain were used to determine face validity and capacity to discriminate neck pain from pain originating in other areas of the body. Results of the study indicated all 20 items were internally consistent ($r = 0.93$). Test-retest reliability was not assessed. The measure was capable of distinguishing neck pain patients from patients with pain located in other areas as well as from healthy volunteers. The NPAD correlated significantly with the OBPQ ($r=0.78$, $P<0.05$) and the Pain and Disability Index ($r=0.80$, $P<0.05$) but not as well with the Beck Depression Inventory ($r=0.52$, $P<0.05$).

Northwick Park Neck Pain Questionnaire

Similar to the NDI in style and domains assessed, the NPNPQ³² was also adapted from the ODQ. The NPQ consists of 10 items, 9 of which assess pain intensity and duration, the effect of neck pain on ADL, work and social involvement, as well as presence of paresthesia or anesthesia. The 10th item elicits the patient's global assessment of the progression of his or her complaint. The NPNPQ and the NDI assess 6 of the same domains with wording that is very similar due to their common origin. Using a Likert scale with scores from 0 to 4, the patient is instructed to mark 1 of 5 boxes to indicate the statement that most closely describes their problem. To accommodate a European patient base, the question regarding driving may be omitted if not applicable. The NPNPQ is scored by summing the domain scores, dividing by 36 (or 32 if the driving-related question is eliminated) and converting to a percentage, 0% representing absence of disability and 100% indicating severe disability.

The NPNPQ was administered 4 times over 3 months to a total of 44 patients (31 female, 13 male; mean age 48 years) presenting with mechanical neck pain. Only 35 patients completed the final assessment phase. Statistical analysis of the data suggests that NPNPQ scores are internally consistent, mean scores tending to rise with increased pain intensity. The NPQ also demonstrates short-term repeatability ($\kappa = 0.53 - 0.76$). Regarding responsiveness, the mean scores decreased at each assessment but significant overall change was not demonstrated. There was significant correlation between the patient's global assessment and pain intensity as well as changes in total NPNPQ scores.

Patient-Specific Functional Scale

More so than any of the other outcome measures discussed within this review, the Patient-Specific Functional Scale (PSFS)¹⁹ (Appendix D) allows the patient to determine which of 5 activities are affected most by their neck condition and then monitors changes in performance capabilities. Additionally, there are 2 questions modeled after pain questions on the SF-36 that assess pain intensity and activity limitation due to pain. Scoring for the PSFS is based on an 11-point numerical rating scale with 0 = unable to perform activity and 10 = able to perform activity at pre-injury level. The pain intensity scale is graded as 0 = no pain and 10 = pain as

TABLE 3
SUBJECTIVE OUTCOME ASSESSMENT FOR USE WITH CERVICAL SPINE PATHOLOGY

OUTCOME ASSESSMENT	CATEGORY	DOMAINS/ DIMENSIONS; NUMBER OF QUESTIONS	TYPE OF SCALE	METHOD OF SCORING	TIME TO COMPLETE TIME TO SCORE
BOURNEMOUTH NECK QUESTIONNAIRE ²⁹	REGION-SPECIFIC	7 QUESTIONS ASSESSING, PAIN SEVERITY, DISABILITY, AFFECTIVE AND COGNITIVE ASPECTS OF NECK PAIN	ORDINAL: 11-POINT NUMERICAL RATING SCALE FROM 0 TO 10		
CERVICAL SPINE OUTCOMES QUESTIONNAIRE ¹⁶	REGION-SPECIFIC	33 QUESTIONS ASSESSING PAIN SEVERITY FOR NECK, SHOULDER AND ARM, ADL LIMITATIONS, PSYCHOSOCIAL COMPONENTS, PHYSICAL SYMPTOMS HEALTH CARE UTILIZATION AND DRUG USE.	NOMINAL & ORDINAL: Q 1–12: 6 COMPARISONS OF RELATIVE MAGNITUDE PER QUESTION Q 13–25: 4 COMPARISONS OF RELATIVE MAGNITUDE PER QUESTION Q 26–30, 32, 33: YES/NO Q 31: FREQ. OF DRUG USE	RESPONSES TO INDIVIDUAL QUESTIONS CONVERTED TO NUMERICAL VALUES. SCORES WITHIN EACH OF THE 6 DOMAINS WERE COMBINED TO CREATE A COMPOSITE DOMAIN SCORE (0–100);	INFORMATION NOT REPORTED
COPENHAGEN NECK FUNCTIONAL DISABILITY SCALE ³¹	REGION-SPECIFIC	15 QUESTIONS ASSESSING ADLS, AFFECTIVE DIMENSIONS	NOMINAL: “YES, OCCASIONALLY, NO”	Q 1–5: YES = 0; OCCASIONALLY = 1; NO = 2 Q 6–15: YES = 2; OCCASIONALLY = 1; NO = 2 MAXIMUM SCORE, INDICATING EXTREME DISABILITY = 30	INFORMATION NOT REPORTED
DISABILITY RATING INDEX ³⁴	GENERAL	12 QUESTIONS ASSESSING ADLS, WORK CAPACITY, PHYSICAL ACTIVITY	ORDINAL: 100-MM VISUAL ANALOG SCALE	MEASURE DISTANCE BETWEEN 0 AND PATIENT’S MARK. DISTANCES CAN BE ADDED AND THE MEAN SCORE USED AS THE INDEX; DISTANCES CAN BE PLOTTED INDIVIDUALLY TO GIVE A DETAILED PROFILE.	2–4 MINUTES; 1–2 MINUTES
EXTENDED ABERDEEN SPINE PAIN SCALE ³⁶	REGION-SPECIFIC	28 QUESTIONS OVERALL; 18 QUESTIONS APPLICABLE TO THE CERVICAL SPINE ASSESS PAIN FREQUENCY & LOCATION; PROVOCATIVE FACTORS, MEDICATION USE AND ADLS	ORDINAL: DESCRIPTORS AND NUMBER OF POSSIBLE RESPONSES VARY	INFORMATION NOT PROVIDED	>5 MINUTES; INFORMATION NOT PROVIDED REGARDING SCORING TIMES

TABLE 3
CONTINUED

OUTCOME ASSESSMENT	CATEGORY	DOMAINS/ DIMENSIONS; NUMBER OF QUESTIONS	TYPE OF SCALE	METHOD OF SCORING	TIME TO COMPLETE TIME TO SCORE
FUNCTIONAL RATING INDEX ³⁰	GENERAL	10 SECTIONS ASSESSING PAIN INTENSITY, FREQUENCY AND ADLS	ORDINAL: 10 CM VISUAL ANALOG SCALE DIVIDED INTO 4 EQUAL SECTIONS WITH 5 RELATIVE MAGNITUDE DESCRIPTORS	POSSIBLE RESPONSES GRADED ON A 0–4 SCALE; TOTAL SCORE IS DIVIDED BY 40 AND CONVERTED TO PERCENTAGE. GREATER SCORES INDICATE MORE SEVERE DYSFUNCTION AND/OR PAIN.	AVERAGE OF 64 SECONDS TO COMPLETE; AVERAGE OF 14 SECONDS TO SCORE
NECK DISABILITY INDEX ¹⁵	REGION-SPECIFIC	10 SECTIONS ASSESSING PAIN INTENSITY, SYMPTOMS, ADLS AND CONCENTRATION	ORDINAL: 6 COMPARISONS OF RELATIVE MAGNITUDE PER SECTION	ITEMS CORED FROM 0 TO 5; TOTAL SCALE SCORE: TOTAL INDIVIDUAL ITEM SCORES, MULTIPLY BY TWO AND CONVERTING TO A PERCENTAGE.	INFORMATION NOT REPORTED
NECK PAIN AND DISABILITY VISUAL ANALOG SCALE ³⁵	REGION-SPECIFIC	20 QUESTIONS ASSESSING PAIN INTENSITY, ADLS, AFFECTIVE COMPONENTS, CERVICAL MOTION	ORDINAL: 10 CM VISUAL ANALOG SCALE		LESS THAN 5 MINUTES; SCORING TIME NOT PROVIDED
NORTHWICK PARK NECK PAIN QUESTIONNAIRE ³²	REGION-SPECIFIC	10 SECTIONS ASSESSING PAIN INTENSITY, SYMPTOMS, ADLS AND PROGRESSION	ORDINAL: 5 COMPARISONS OF RELATIVE MAGNITUDE PER SECTION	ANSWERS SCORED 0–4 AND TOTALED	NOT INDICATED
PATIENT-SPECIFIC FUNCTIONAL SCALE ¹⁹	PATIENT-SPECIFIC	3 SECTIONS THAT ASSESS PAIN LIMITATION, PAIN INTENSITY, AND 3 ADLS SELECTED BY THE PATIENT	ORDINAL: 11-POINT NUMERICAL RATING SCALE FROM 0 TO 10	0 = UNABLE TO PERFORM ACTIVITY; 10 = ABLE TO PERFORM ACTIVITY AT PREINJURY LEVEL. PAIN INTENSITY SCALE GRADED AS 0 = NO PAIN AND 10 = PAIN AS BAD AS IT CAN BE. ANSWERS TOTALED AND CONVERTED TO PERCENTAGE.	NOT INDICATED
WHIPLASH DISABILITY QUESTIONNAIRE ³³	CONDITION-SPECIFIC	13 QUESTIONS ASSESSING SYMPTOM INTERFERENCE WITH ADLS AND AFFECTIVE PARAMETERS	ORDINAL: 11-POINT NUMERICAL RATING SCALE FROM 0 TO 10	0 = NOT AT ALL AND 10 = INABILITY TO PERFORM A FUNCTION OR EXTREMELY NEGATIVE EFFECT. SCORES TOTALED.	NOT INDICATED

TABLE 4.
SUBJECTIVE OUTCOME ASSESSMENT FOR USE WITH CERVICAL SPINE PATHOLOGY—RELIABILITY,
VALIDITY, AND RESPONSIVENESS

OUTCOME ASSESSMENT	PARTICIPANTS, CONDITIONS	RELIABILITY	VALIDITY	RESPONSIVENESS TO CHANGE
BOURNEMOUTH NECK QUESTIONNAIRE ²⁹	102 PATIENTS (64% FEMALE) WITH NONSPECIFIC NECK PAIN; MEAN AGE = 45.4 YRS	CRONBACH'S ALPHA = 0.87, 0.91, 0.93 ON (3) ADMINISTRATIONS; ICC FOR EACH OF 7 ITEMS RANGED FROM 0.50 TO 0.63; ICC OF TOTAL SCORE = 0.65	PEARSON'S R (PRE-TREATMENT) = 0.63 PEARSON'S R (POST-TREATMENT) = 0.48 STATISTICAL CORRELATION WITH NDI	PEARSON'S R FOR ALL 7 ITEMS RANGED FROM 0.42 (ITEM 5) TO 0.82 (ITEM 2); EFFECT SIZE (KAZIZ AND COHEN METHODS) = 1.67 (BASELINE SCORES) AND 1.43 (CHANGE SCORES)
CERVICAL SPINE OUTCOMES QUESTIONNAIRE ¹⁶	POOL OF 215 PATIENTS; THOSE SELECTED FOR ANTERIOR CERVICAL SPINE FUSION WERE CONSIDERED ELIGIBLE BUT NO DATA AVAILABLE FOR EXACT NUMBER OF PARTICIPANTS	TEST-RETEST (3 MONTHS & 6 MONTHS) ICC SCORES: 0.80–0.86 FOR ALL COMPOSITE MEASURES, EXCEPT HEALTH CARE UTILIZATION (0.75)	CONSTRUCT VALIDITY: PAIN, FUNCTIONAL DISABILITY, PSYCHOLOGICAL DISTRESS AND PHYSICAL SYMPTOM SCORES WERE HIGHLY CORRELATED WITH SF-36 AND ODI SCORES.	MODERATELY RESPONSIVE (EXCEPT FOR HEALTH CARE UTILIZATION QUESTION)
COPENHAGEN NECK FUNCTIONAL DISABILITY SCALE ³¹	162 PATIENTS, 39 OF WHICH HAD A HX OF CERVICAL SURGERY, 117 MECHANICAL NECK PAIN AND 21 SEEKING TREATMENT	SHORT TERM TEST-RETEST: PEARSON'S R = 0.96–0.99 ICC = 0.95–0.99 OVERALL CRONBACH'S ALPHA = 0.897	CONSTRUCT VALIDITY: PAIN AND DISABILITY SCORES CORRELATED WITH SELF-REPORTED PAIN SCORES (PEARSON'S R = 0.89) AND TO DOCTOR'S AND PATIENT'S GLOBAL ASSESSMENT SCORES (SPEARMAN RANK CORRELATION COEFFICIENT R = 0.56 AND R = 0.89 RESPECTIVELY).	SELF-REPORTED PAIN SCORES WERE CORRELATED TO DISABILITY SCORES AT THE CONCLUSION OF TREATMENT, AT 4-MONTH FOLLOW-UP AND AT 12-MONTH FOLLOW-UP, INDICATING STRONG CORRELATION.
DISABILITY RATING INDEX ³⁴	1458 PARTICIPANTS TOTAL DIVIDED INTO 10 COHORTS; (1) NECK/SHOULDER PAIN GROUP (N = 127; 45 FEMALES, 82 MALES) AND (5) NECK/SHOULDER/LOW BACK PAIN GROUPS (N = 176; 115 FEMALES, 61 MALES). MEDIAN AGE = 40.5 YEARS. PATIENTS SELECTED FROM INDUSTRIAL HEALTH CARE CENTERS, INSURANCE OFFICES, AND REHABILITATION UNITS.	1-DAY TEST-RETEST RELIABILITY: PEARSON'S CORRELATION = 0.95 OVERALL 3-DAY TEST-RETEST RELIABILITY: PEARSON'S CORRELATION = 0.92 OVERALL INTER-RATER RELIABILITY: 0.99 INTRA-RATER RELIABILITY: 0.98 CRONBACH'S ALPHA = 0.84 (BORDERLINE ACCEPTABLE)	CONSTRUCT VALIDITY: MEAN SCORES FOR HEALTHY GROUP = 0.8–0.9% MEAN SCORES FOR NECK/SHOULDER/LOW BACK PATIENTS = 27–39% MEAN SCORES FOR WHEELCHAIR-BOUND PATIENTS = 85% PEARSON'S ICC (DRI-FSQ) = 0.46 PEARSON'S ICC (DRI-OPDQ) = 0.38 FACE VALIDITY: ROBUST	% SCORES FOR ARTHRITIC PATIENTS: PRE-SURGERY: 52% POST-SURGERY: 36%

bad as it can be. The authors included 31 consecutive neck pain patients (16 female, 15 male; mean age 40.4 years), 2 of which demonstrated radicular findings. All patients were in the acute phase at the time of inclusion.

Test-retest reliability was assessed by comparing initial PSFS scores of a subset of patients who had a poorer prognosis to scores achieved on their second visit. The ICC was 0.92 (lower 95% confidence limit = 0.78) for the average of 3 activities, 0.95 (lower

TABLE 4.
CONTINUED

OUTCOME ASSESSMENT	PARTICIPANTS, CONDITIONS	RELIABILITY	VALIDITY	RESPONSIVENESS TO CHANGE
EXTENDED ABERDEEN SPINE PAIN SCALE ³⁶	512 CONSECUTIVE PATIENTS (64% WOMEN, 36% MEN; MEAN AGE: 44.6 YEARS) PRESENTING TO A GENERAL PRACTICE OR THOSE PARTICIPATING IN AN RCT OF OSTEOPATHY VS. GENERAL PRACTICE CARE; 388 WITH PAIN IN A SINGLE AREA; 249 NECK SCORES	TEST-RETEST: MEAN DIFFERENCE IN SCORES OF 4.7, SD OF 9.2	FACE VALIDITY: AGREED ON BY 10 PATIENTS AND 15 PROFESSIONALS; INTERNAL CONSISTENCY: CRONBACH'S ALPHA = 0.9 FOR NECK; CONSTRUCT VALIDITY: SCORES IMPROVED BY ONE STANDARD DEVIATION OR GREATER—ONLY MEAN CHANGES IN SCORES GIVEN; CRITERION VALIDITY: PEARSON'S CORRELATION COEFFICIENT WERE HIGHER WITH PHYSICAL SCORES THAN WITH MENTAL SCORES—ONLY MEAN CHANGES IN SCORES GIVEN.	RESPONSES SHOWED LARGE CHANGES WHEN HEALTH STATUS IMPROVED AND SMALLER CHANGES WHEN HEALTH STATUS DETERIORATED.
FUNCTIONAL RATING INDEX ³⁰	139 PATIENTS RECEIVING CHIROPRACTIC TREATMENT FOR SPINAL CONDITIONS; 60% FEMALE, MEAN AGE 41 YEARS, 89.3% CAUCASIAN. SUBJECTS ASSIGNED TO 1 OF 4 COHORTS.	TEST-RETEST RELIABILITY (N = 51): ICC = 0.99 CRONBACH'S ALPHA = 0.92 SPEARMAN'S RHO = 0.54–0.82 ALL ITEMS MODERATELY CORRELATED	CONSTRUCT VALIDITY (N = 24): WITH SF-12 PCS: PEARSON'S PRODUCT MOMENT = 0.76) WITH DRI: PEARSON'S = 0.76 CONTENT VALIDITY: FOUND TO BE APPROPRIATE.	STANDARD RESPONSE MEAN (N = 36): 1.26, INDICATING LARGE EFFECT SIZE.
NECK DISABILITY INDEX ¹⁵	48 SUBJECTS (17 MALES); AVERAGE AGE = 35 FOR FEMALES AND 41 FOR MALES	TEST-RETEST RELIABILITY: CORRELATION COEFFICIENT = 0.89 ALPHA COEFFICIENTS FOR INTERNAL CONSISTENCY (TOTAL = 0.80)	CONCURRENT VALIDITY WITH NDI: 0.60	NOT DISCUSSED
NECK PAIN AND DISABILITY VISUAL ANALOG SCALE ³⁵	100 CONSECUTIVE PATIENTS (39 MALES) WITH AVERAGE AGE OF 44.26 YEARS; 82% REPORTED CHRONIC PAIN (>3 MONTHS DURATION)	INTERNAL CONSISTENCY ALPHA R = 0.93; ITEMS SHOWED STRONG CORRELATION WITH TOTAL SCORE	FACE VALIDITY DETERMINED; CORRELATED WITH ODQ (0.78) AND PDI 90.80	NOT DISCUSSED
NORTHWICK PARK NECK PAIN QUESTIONNAIRE ³²	44 PATIENTS WITH NECK PAIN WHO ATTENDED A RHEUMATOLOGY CLINIC OVER A 27-WEEK PERIOD; 31 WOMEN; AVERAGE AGE OF PATIENTS = 48 YEARS	SHORT-TERM REPEATABILITY: R = 0.84, K = 0.62; INTERNAL CONSISTENCY DEMONSTRATED	NOT DETERMINED	QUESTIONABLE

95% confidence limit = 0.86) for pain limitation and 0.81 (lower 95% confidence limit = 0.56) for pain intensity. The NDI in this study demonstrated an ICC of 0.89. The authors chose a unique method to

assess construct validity, namely comparing PSFS scores to the clinician's estimate of the patient's prognosis rated on a 5-point scale. Specifically, the PSFS change scores were compared to the clinician's

TABLE 4.
CONTINUED

OUTCOME ASSESSMENT	PARTICIPANTS, CONDITIONS	RELIABILITY	VALIDITY	RESPONSIVENESS TO CHANGE
PATIENT-SPECIFIC FUNCTIONAL SCALE ¹⁹	31 CONSECUTIVE PATIENTS WITH NECK PAIN (16 FEMALE) WITH AVERAGE AGE OF 40.4 YEARS; NONE OF THE PATIENTS WERE CONSIDERED TO BE CHRONIC PAIN PATIENTS.	R = 0.92	R = 0.73–0.83 (COMPARED WITH NDI) R = 0.52–0.64 (COMPARED TO PROGNOSTIC RATING)	SENSITIVITY TO CHANGE: R = 0.79–0.83 COMPARED WITH NDI; R = 0.46–0.53 COMPARED TO PROGNOSIS RATING
WHIPLASH DISABILITY QUESTIONNAIRE ³³	101 PARTICIPANTS (81% FEMALE) WITH MEAN AGE OF 40.3 YEARS. MEDIAN OF 18 MONTHS ELAPSED TIME SINCE WHIPLASH INJURY	INTERNAL CONSISTENCY: CRONBACH'S ALPHA = 0.96	FACE VALIDITY DETERMINED	NOT DISCUSSED

prognostic rating of change. Correlation existed between PSFS activity scores and those of the NDI at intake ($r = 0.82$) and discharge ($r = 0.83$) but was lower when comparing the PSFS to the prognostic rating ($r = 0.64$)

Responsiveness was assessed by comparing change scores of the PSFS to change scores of both the NDI (Pearson's $r = 0.83$) and the prognostic rating (Spearman's $r = 0.53$). The change scores between the NDI and prognostic rating were similar (Spearman's $r = 0.54$). The authors note that due to specificity of patient-defined items, comparison among patients is nearly impossible.

Whiplash Disability Questionnaire

Following a review of the literature, Pinfold and colleagues determined that a condition-specific disability outcome measure for whiplash associated disorders did not exist and could be of benefit as these injuries frequently occur following motor vehicle accidents. Based on the work of Hoving et al,⁴⁷ the authors developed the Whiplash Disability Questionnaire (WDQ),³³ a 13-item measure of pain, ADL, recreational and social activities and 4 questions specific to the affective component of pain. Included domains were also selected from the NDI as well as volunteered by patients. Patients are asked to rate the effect of their whiplash injury and symptoms using a 0 to 10 numerical rating scale with 0 = not at all and 10 = inability to perform a function or extremely negative effect.

Patients recruited for the validation study included 82 females and 19 males with a mean age of 40.3 years. All patients had experienced a whiplash associated disorder and no restriction was placed on elapsed time since injury; median time was 18 months and average time was 48 months. All patients were seeking treatment from private practice physiotherapists.

The questionnaire was given to content experts for comment regarding relevance of domains and ease of application. Based on feedback, the questionnaire demonstrated face validity. Internal consistency was excellent (Cronbach's $\alpha = 0.96$). Content validity was acceptable and no substantial floor (12%) or ceiling (4%) effects were identified.

DISCUSSION:

With a shared goal to assess subjective parameters of cervical spine pain, the outcome assessment tools included in this review have much in common. The specifics of each of the eleven outcome measures, including types of domains, method of scoring, and data on reliability, validity and responsiveness are listed in Tables 3 and 4.

The most frequently assessed domains were related to personal care, ADL, work, recreation/leisure activities, pain intensity, sleeping, driving, lifting/carrying capacity, social activities, concentration and reading (Table 5). Interestingly, only 3 measures included questions about medication use and only 1 assessed the frequency of pain, a parameter that

TABLE 5.
FREQUENCY OF SPECIFIC DOMAINS ASSESSED BY EACH OUTCOME ASSESSMENT INSTRUMENT

DOMAIN	NDI	NPQ	DRI	PSFS	CNFDS	NPAD	FRI	EABPS	BNQ	CSOQ	WDQ	TOTAL
PERSONAL CARE/ADLS	✓		✓	✓		✓	✓	✓	✓	✓	✓	9
WORK	✓	✓	✓			✓	✓	✓	✓	✓	✓	9
RECREATION/ LEISURE-WALKING, STANDING, RUNNING	✓		✓		✓	✓	✓	✓	✓	✓	✓	9
PAIN INTENSITY	✓	✓		✓		✓	✓		✓	✓	✓	8
SLEEPING	✓	✓			✓	✓	✓	✓		✓	✓	8
DRIVING	✓	✓				✓	✓	✓			✓	7
LIFTING/CARRYING	✓	✓	✓		✓	✓			✓			6
SOCAL ACTIVITIES		✓			✓	✓			✓	✓	✓	6
CONCENTRATION	✓				✓	✓			✓		✓	5
READING	✓	✓			✓			✓	✓			5
MEDICATION USE						✓		✓		✓		3
HEADACHES	✓				✓					✓		3
STAY IN BED MORE					✓			✓		✓		3
ANXIETY LEVEL									✓	✓	✓	3
DEPRESSION									✓	✓	✓	3
PRESENCE OF PINS/ NEEDLES OR NUMBNESS		✓								✓		2
STANDING BENT OVER SINK			✓		✓							2
EMOTIONAL RELATIONSHIPS					✓	✓						2
PAIN INFLUENCE ON FUTURE					✓	✓						2
PAIN FREQUENCY							✓	✓				2
SEX LIFE								✓		✓		2
LOW ENERGY LEVEL										✓	✓	2
GLOBAL RATING		✓										1
CLIMBING STAIRS			✓									1
SITTING			✓									1
MAKING BED			✓									1
PAIN LIMITATION				✓								1
STAY AT HOME MORE					✓							1
DURATION OF SYMPTOMS		✓										1
WATCHING T.V.		✓										1
ACTIVE MOTIONS						✓						1
PROVOCATIVE FACTORS								✓				1
LOCATION OF SYMPTOMS								✓				1
CONTROL									✓			1
WORRIED										✓		1
FELT SICK OR UNWELL										✓		1
ANGER											✓	1

NECK DISABILITY INDEX (NDI); NORTHWICK PARK PAIN QUESTIONNAIRE (NPQ); DISABILITY RATING INDEX (DRI); PATIENT-SPECIFIC FUNCTIONAL SCALE (PSFS); COPENHAGEN NECK FUNCTIONAL DISABILITY SCALE (CNFDS); NECK PAIN AND DISABILITY VISUAL ANALOG SCALE (NPAD); FUNCTIONAL RATING INDEX (FRI); EXTENDED ABERDEEN SPINE PAIN SCALES (EABPS); BOURNEMOUTH NECK QUESTIONNAIRE (BNQ); CERVICAL SPINE OUTCOMES QUESTIONNAIRE (CSOQ); WHIPLASH DISABILITY QUESTIONNAIRE (WDQ)

may be as pertinent to the pain experience as is intensity. Concerning scale types used, only the DRI used a true VAS format on which to record patient responses. Both the FRI and the NPAD used segmented visual analog scales with descriptors at the end points (NPAD) and at each division (FRI). The BNQ, PSFS and the Whiplash Disability Questionnaire each used an, 11-point numerical rating scale. The Extended Aberdeen Spine Pain Scale, the NDI and the NPNQ employed a Likert scale with 5 to 6 response options in sentence format listed in increasing order of disability. Similarly, the CSOQ makes use of 1 word descriptors ranging from *none*

to *excruciating* while the CNFDS uses 3 patient response options: *yes*, *occasionally* and *no*.

Each of the 11 outcome measures focuses on a selected group of domains. None of them assesses all domains completely and this is just as well for such an instrument would potentially be cumbersome and time-consuming for patients and clinicians alike. The instruments were tested on a varied patient base, including pre- and post-surgical patients, as well as acute and chronic mechanical neck pain patients. Subsequent testing of the majority of these instruments on specific patient populations has not

been undertaken. An exception to this statement occurs with the NDI,^{18,46,47} the NPAD,⁴⁸ and the NPNPQ⁴⁷ which have been assessed beyond their initial validation studies. Additionally, the NPNPQ,^{38,39,41} the NDI^{40,41} and the NPAD⁴¹ have been translated to languages other than English and the authors determined each to be a reliable and valid measure within the targeted population.

It is unknown whether these outcome assessment tools would demonstrate the same levels of reliability and validity when applied to populations different than those used in the initial testing process. One would surmise that parameters considered significant to athletes may not be similar to those of sedentary retirees. Likewise, many of these studies were conducted on largely Caucasian patients and the data may not be generalized to populations of other ethnic or cultural backgrounds. Additionally, it remains undetermined whether there is a need for measures specific to acute and chronic conditions. The responsiveness data given in the studies included in this review suggest that each can discern clinically relevant change at acceptable levels. However, a cohort study does not exist wherein these populations are specifically compared.

Not included in this review is the North American Spine Society (NASS) Cervical Spine Outcome Assessment tool. According to NASS (personal communications with K. Hintz, 3 Sept 2004), the instrument is currently undergoing revision and is unavailable for review. Also not included is the cervical non-organic signs⁴⁹ tool as it is an objective measure of physical signs than may present in patients with abnormal illness behavior.

CONCLUSIONS

Given that each of the subjective outcome assessment tools presented in this review demonstrated acceptable levels of reliability and validity, and varying degrees of responsiveness, the choice of which to use becomes a matter of preference based on the particular patient population each clinician treats, the ease of use for the patient and ease of scoring for the clinician, and the domains that are most relevant to a particular patient. It is quite likely that pain intensity, ADL and work restrictions may be the most important parameters for one patient while for another patient pain may have much more of an impact on affective domains. Therefore, the clinician must take time to determine which

areas are most significant for each patient and choose the appropriate instrument for the task. In this regard, developing a gold standard for subjective outcome measures applicable to the cervical spine, while desirable for comparing the results of treatment between studies and between patients, may be a difficult task indeed given the diversity of biopsychosocial influences on each patient's experience. A more acceptable alternative, as suggested by Bombardier and colleagues,⁵⁰ would be to agree on a standard set of outcome measures that would allow comparison of treatment effects across studies.

REFERENCES

1. Côté, P, Cassidy JD, Carroll L. The Saskatchewan Health and Back Pain Survey: the prevalence of neck pain and related disability in Saskatchewan adults. *Spine* 1998;23:1689-98.
2. Mäkelä M, Heliövaara M, Sievers K et al. Prevalence, determinants, and consequences of chronic neck pain in Finland. *Am J Epidemiol* 1991;134:1356-67.
3. Barry M, Jenner JR. Pain in neck, shoulder and arm. *Br Med J* 1995; 310:183-6.
4. Gross AR, Hondras MA, Aker PD, et al. Manual therapy for mechanical neck disorders. *Cochrane Library* [CD-ROM] 2000;2.
5. Hallgren RC, Greenman PE, Rechten JJ. Atrophy of suboccipital muscles in patients with chronic pain: a pilot study. *J Am Osteopath Assoc* 1995;12:1032-9.
6. Davis CK. The federal role in changing health care financing. Part 1: national programs and health financing problems. *Nurs Econ* 1983; 1:10-17.
7. Mowinski Jennings B, Staggers N. The language of outcomes. *J Rehabil Outcomes Meas* 1999;3:59-64.
8. Roper WL, Hackbarth GM. Commentary. HCFA's agenda for promoting high-quality care. *Health Affairs* 1988;1:91-98.
9. Wennberg JE. Improving the medical decision-making process. *Health Aff* 1988;7:99-106.
10. Hutchinson A, Mc Coll E, Christie M, Riccalton C. Health outcomes for primary and out-patient care: an editorial overview. In: Hutchinson A. Health outcome measures in primary and outpatient care. The Netherlands: Harwood Academic Publishers; 1996. p. 11.
11. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I: conceptual framework and item selection. *Med Care* 1992;30:473-83.
12. Nelson E, Wasson J, Kirk J, et al. Assessment of function in routine clinical practice: description of the COOP Chart method and preliminary findings. *J Chronic Dis* 1987;40:55S-69S.
13. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders. Summary and general recommendations. *Spine* 2000;25:3100-3103.
14. Fairbank JCT, Couper J, Davies JB, O'Brian JP. The Oswestry low back pain disability index. *Physiotherapy* 1980;66:271-3.
15. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14:409-15.
16. BenDebba M, Heller J, Ducker TB, Eisinger JM. Cervical spine outcomes questionnaire. Its development and psychometric properties. *Spine* 2002;27:2116-2124.
17. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:S217-32.
18. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther* 1998;78:951-63.
19. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 1998;27:331-8.
20. Bolton JE, Breen AC. The Bournemouth questionnaire: a short-form

- comprehensive outcome measure. I: psychometric properties in back pain patients. *J Manipulative Physiol Ther* 1999;22:503–10.
21. Nagi SZ. Some conceptual issues in disability and rehabilitation. In: Sussman MB, editor. *Sociology and rehabilitation*. Washington, DC: American Sociological Association; 1965. p. 100–13.
 22. Hansen DT, Mior S, Mootz RD. Why outcomes? Why now? In: Yeomans SG, editor. *The clinical application of outcomes assessment*. Stamford, CT: Appleton & Lange; 2000. p. 9–12.
 23. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *Br J Med* 1997;314:572.
 24. Keating JC. *Toward a philosophy of the science of chiropractic: a primer for clinicians*. Stockton, CA: Stockton Foundation for Chiropractic Research; 1992. p. 130.
 25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–415.
 26. Wyrwich K, Nienaber N, Tierney W, Wolinsky F. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–78.
 27. Stratford PW, Binkley J, Solomon P, et al. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther* 1996;74:359–65.
 28. Cohen J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press; 1977. p. 1–27.
 29. Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II: psychometric properties in neck pain patients. *J Manipulative Physiol Ther* 2002;25:141–8.
 30. Feise RJ, Menke MJ. Functional rating index: a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26:78–86.
 31. Jordan A, Manniche C, Mosdal C, Hindsberger C. The Copenhagen neck functional disability scale: a study of reliability and validity. *J Manipulative Physiol Ther* 1998;21:520–7.
 32. Leak AM, Cooper J, Dyer S, et al. The Northwick Park neck pain questionnaire, devised to measure neck pain and disability. *Br J Rheumatol* 1994;33:469–74.
 33. Pinfold M, Niere KR, O'Leary EF, et al. Validity and internal consistency of a whiplash-specific disability measure. *Spine* 2004;29:263–8.
 34. Salen BA, Spangfort EV, Nygren AL, Nordemar R. The disability rating index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol* 1994;47:1423–34.
 35. Wheeler AH, Goolkasian P, Baird AC, Darden BV. Development of the neck pain and disability scale: item analysis, face, and criterion-related validity. *Spine* 1999;24:1290–4.
 36. Williams NH, Wilkinson C, Russell IT. Extending the Aberdeen back pain scale to include the whole spine: a set of outcome measures for the neck, upper and lower back. *Pain* 2001;94:261–274.
 37. Pietrobon R, Coeytaux RR, Carey S, et al. Standard scales for measurement of functional outcome for cervical pain or dysfunction. A systematic review. *Spine* 2002;27:515–522.
 38. Chiu TT, Lam TH, Hedley AJ. Subjective health measure used on Chinese patients with neck pain in Hong Kong. *Spine* 2001;26:1884–9.
 39. Gonzalez T, Balsa A, Sainz de Murieta J, et al. Spanish version of the Northwick Park neck pain questionnaire: reliability and validity. *Clin Exp Rheumatol* 2001;19:41–6.
 40. Helmersen Ackelman B, Lindgren U. Validity and reliability of a modified version of the neck disability index. *J Rehab Med* 2002; 34:284–7.
 41. Wlodyka-Demaille S, Poiraudreau S, Catanzariti JF, et al. French translation and validation of 3 functional disability scales for neck pain. *Arch Phys Med Rehabil* 2002;83:376–82.
 42. Stratford P, Solomon P, Binkley J, et al. Sensitivity of sickness impact profile items to measure change over time in a low-back pain patient group. *Spine* 1993;18:1723–7.
 43. Melzak R, Torgerson WS. On the language of pain. *Anesthesiology* 1971;34:50–9.
 44. Ruta DA, Garratt AM, Wardlaw D, Russell IT. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994;19:1887–1896.
 45. Melzak R. The McGill pain questionnaire: major properties and scoring methods. *Pain* 1975;1:275–99.
 46. Hains F, Waalen J, Mior S. Psychometric properties of the neck disability index. *J Manipulative Physiol Ther* 1998;21:75–80.
 47. Hoving JL, O'Leary EF, Niere KR, et al. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 2003;102:273–281.
 48. Goolkasian P, Wheeler AH, Gretz SS. The neck pain and disability scale: test-retest reliability and construct validity. *Clin J Pain* 2002;18: 245–250.
 49. 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.
 50. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheum* 2001;28:431–8.

Appendix A - Bournemouth Neck Questionnaire

The following scales have been designed to find out about your neck pain and how it is affecting you. Please answer ALL the scales by circling ONE number on EACH scale that best describes how you feel:

1. Over the past week, on average how would you rate your neck pain?

No pain										Worst pain possible
0	1	2	3	4	5	6	7	8	9	10

2. Over the past week, how much has your neck pain interfered with your daily activities (housework, washing, dressing, lifting, reading, driving)?

No interference										Unable to carry out activities
0	1	2	3	4	5	6	7	8	9	10

3. Over the past week, how much has your neck pain interfered with your ability to take part in recreational, social, and family activities?

No interference										Unable to carry out activities
0	1	2	3	4	5	6	7	8	9	10

4. Over the past week, how anxious (tense, uptight, irritable, difficulty in concentrating/relaxing) have you been feeling?

Not at all anxious										Extremely anxious
0	1	2	3	4	5	6	7	8	9	10

5. Over the past week, how depressed (down-in-the-dumps, sad, in low spirits, pessimistic, unhappy) have you been feeling?

Not at all depressed										Extremely depressed
0	1	2	3	4	5	6	7	8	9	10

6. Over the past week, how have you felt your work (both inside and outside the home) has affected (or would affect) your neck pain?

Have made it worse										Have made it much worse
0	1	2	3	4	5	6	7	8	9	10

7. Over the past week, how much have you been able to control (reduce/help) your neck pain on your own

Completely control it										No control whatsoever
0	1	2	3	4	5	6	7	8	9	10

The Bournemouth Neck Questionnaire. Reprinted from Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II: psychometric properties in neck pain patients. J Manipulative Physiol Ther 2002;25:141-8, with permission from the National University of Health Sciences.

Appendix B - Disability Rating Index

Date: _____ Name: _____

Without difficulty	Example	Not at all
With some difficulty	--- With difficulty ---	With great difficulty

How do you manage the following activities?
After each question, please mark ONE POINT on the line. Please answer all questions

Dressing (without help)	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Out-door walks	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Climbing stairs	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Sitting longer time	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Standing bent over a sink	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Carrying a bag	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Making a bed	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Running	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Light work	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Heavy work	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Lifting heavy objects	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>
Participating in exercise/sports . . .	<div style="position: absolute; left: -5px; top: -5px; right: -5px; height: 10px; border: 1px solid black;"></div>

Disability Rating Index. Reprinted from Salen BA, Spangfort EV, Nygren AL, Nordemar R. The disability rating index: an instrument for the assessment of disability in clinical settings. J Clin Epidemiol 1994;47:1423-34, with permission from Elsevier.

Appendix C - Neck Disability Index

PATIENT

NAME: _____

FILE #: _____ DATE: _____

This questionnaire has been designed to give the doctor information as to how your neck pain has affected your ability to manage in everyday life. Please answer every section and mark in each section only ONE box which applies to you. We realize you may consider that two of the statements in any one section relate to you, but please just mark the box which most closely describes your problem.

Section 1 – Neck Pain Intensity

- ☐ I have no pain at the moment.
- ☐ The pain is very mild at the moment.
- ☐ The pain is moderate at the moment.
- ☐ The pain is fairly severe at the moment.
- ☐ The pain is very severe at the moment.
- ☐ The pain is the worst imaginable at the moment.

Section 2 – Personal Care (Washing, Dressing, etc.)

- ☐ I can look after myself normally without causing extra pain.
- ☐ I can look after myself normally but it causes extra pain.
- ☐ It is painful to look after myself and I am slow and careful.
- ☐ I need some help but manage most of my personal care.
- ☐ I need help every day in most aspects of self care.
- ☐ I do not get dressed, I wash with difficulty and stay in bed.

Section 3 – Lifting

- ☐ I can lift heavy weights without extra neck pain.
- ☐ I can lift heavy weights but it gives extra neck pain.
- ☐ Neck pain prevents me from lifting heavy weights off the floor, but I can manage if they are conveniently positioned, for example on a table.
- ☐ Neck pain prevents me from lifting heavy weights, but I can manage to lift medium weights if they are conveniently positioned.
- ☐ I can lift very light weights.
- ☐ I cannot lift or carry anything at all.

Section 4 – Reading

- ☐ I can read as much as I want, with no pain in my neck.
- ☐ I can read as much as I want, with slight pain in my neck.
- ☐ I can read as much as I want, with moderate pain in my neck.
- ☐ I cannot read as much as I want because of moderate pain in my neck.
- ☐ I can hardly read at all because of severe pain in my neck.
- ☐ I cannot read at all because of the pain in my neck.

Section 5 – Headaches

- ☐ I have no headaches at all.
- ☐ I have slight headaches which come infrequently.
- ☐ I have moderate headaches which come infrequently.
- ☐ I have moderate headaches which come frequently.
- ☐ I have severe headaches which come frequently.
- ☐ I have headaches almost all the time.

Section 6 – Concentration

- ☐ I can concentrate fully when I want to with no difficulty.
- ☐ I can concentrate fully when I want to with slight difficulty.
- ☐ I have a fair degree of difficulty in concentrating when I want to.
- ☐ I have a lot of difficulty in concentrating when I want to.
- ☐ I have a great, great deal of difficulty in concentrating when I want to.
- ☐ I cannot concentrate at all.

Section 7 Work

- ☐ I can do as much work as I want to.
- ☐ I can only do my usual work, but no more.
- ☐ I can do most of my usual work, but no more.
- ☐ I cannot do my usual work.
- ☐ I can hardly do any work at all.
- ☐ I cannot do any work at all.

Section 8 – Driving

- ☐ I can drive my car without any neck pain at all.
- ☐ I can drive my car as long as I want, with slight pain in my neck.
- ☐ I can drive my car as long as I want, with moderate pain in my neck.
- ☐ I cannot drive my car as long as I want, because of moderate pain in my neck.
- ☐ I can hardly drive at all because of severe pain in my neck.
- ☐ I cannot drive my car at all because of the pain in my neck.

Section 9 – Sleeping

- ☐ I have no trouble sleeping
- ☐ My sleep is barely disturbed (less than 1 hr. sleepless).
- ☐ My sleep is mildly disturbed (1–2 hr. sleepless).
- ☐ My sleep is moderately disturbed (2–3 hr. sleepless).
- ☐ My sleep is greatly disturbed (3–5 hr. sleepless).
- ☐ My sleep is completely disturbed (5–7 hr. sleepless).

Section 10 – Recreation

- ☐ I am able to engage in all my recreational activities, with no neck pain at all.
- ☐ I am able to engage in all my recreational activities, with some pain in my neck.
- ☐ I am able to engage in most, but not all of my usual recreational activities, because of pain in my neck.
- ☐ I am able to engage in a few of my usual recreational activities, because of pain in my neck.
- ☐ I can hardly engage in any recreational activities because of pain in my neck.
- ☐ I cannot engage in any recreational activities at all because of pain in my neck.

Neck Disability Index. Reprinted from Vernon H, Mior S. The neck disability index: a study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–15, with permission from National University of Health Sciences.

Appendix D - Patient-Specific Functional Scale

CLINICIAN TO READ AND FILL IN: Complete at the end of the history and prior to physical

Read at Baseline Assessment

I'm going to ask you to identify up to 3 important activities that you are unable to do or have difficulty with as a result of your problem. Today, are there any activities that you are unable to do or have difficulty with because of your problem? (show scale)

Read at Follow-up Visits

When I assessed you on _____ (slate previous assessment date), you told me that you had difficulty with (read 1, 2, 3 from list). Today, do you still have difficulty with 1 (have patient score each activity): 2 (have patient score each activity): 3 (have patient score each activity).

Scoring scheme (show patient scale):

0	1	2	3	4	5	6	7	8	9	10		
Unable to perform activity											Able to perform activity at pre-injury level	

[illegible]

PAIN QUESTION SET

PAIN LIMITATION

During the past 24 hours, has the pain limited you from performing any of your normal daily activities?

0	1	2	3	4	5	6	7	8	9	10
Unable to perform activity			Able to perform activity at pre-injury level							

PAIN INTENSITY

Over the past 24 hours, how bad has your pain been?
(Point to one number)

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as it can be

Patient-Specific Functional Scale. Reprinted with permission from Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. *J Orthop Sports Phys Ther* 1998;27:331-8.