Detecting Sincerity of Effort: A Summary of Methods and Approaches

Despite the widespread use of methods that are supposed to detect the sincerity of patients' efforts in clinical assessment, little has been written summarizing the literature that addresses the reliability and validity of measurements obtained with these methods. The purpose of this article is to review the literature on the reliability and validity of scores for Waddell's nonorganic signs, descriptions of pain behavior and symptom magnification, coefficients of variation, correlations between musculoskeletal evaluation and function, grip measurements, and the relationship between heart rate and pain intensity. The authors of the articles reviewed conclude that none of these methods have been examined adequately. Some of these methods, such as Waddell's nonorganic signs, were not developed for the purpose of detecting sincerity of effort. Clinicians are encouraged to critically read the literature addressing these methods. With further research, some of the discussed methods may prove useful. Until such research is reported in the peer-reviewed literature, however, clinicians should avoid basing evaluation of sincerity of effort on these tests. Therapists are encouraged, instead, to use a biobehavioral approach to better understand and address the complex factors underlying delayed recovery. [Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. *Phys Ther.* 1998;78:867–888.]

Key Words: *Malingering; Neck and trunk, back; Pain; Reproducibility; Sincerity of effort; Submaximal effort.*

Deborah E Lechner

Sam F Bradbury

Laurence A Bradley

Sincerity of effort, in this article, refers to a patient's conscious motivation to perform optimally during an evaluation.[1-3] A sincere effort is the patient's best or optimal physical performance, whereas an insincere effort is one in which the patient deliberately gives less than a full effort during physical examinations.[1-3] Patients whose efforts are not sincere during evaluation may overuse treatment, may have a prolonged recovery, may have an increased cost of care, or may receive unwarranted disability payments.[1-3] For obvious reasons, there is a keen interest in sincerity of effort from the medicolegal community and from the insurance industry.[1,3,4] Sincerity of effort during evaluation of patients with low back pain (LBP) is frequently questioned when the patient's injury occurs in a work-related environment or when injury-related litigation is pending. Employers, third-party payers, attorneys, and case managers want to know whether the patient with LBP is giving a maximal or best effort during impairment and functional evaluation.

In response to the demand for information on patient participation or cooperation during evaluations, several methods are now widely used by clinicians to evaluate sincerity of effort (Tab. 1).[1-21] These methods are often interpreted by referral sources as measures of sincerity of effort, even when explicit detection of sincerity of effort is not stated in therapists' or clinicians' reports. Do we have evidence in support of the reliability and validity of measurements obtained with these methods for the purpose of detecting sincerity of effort? Are we relying, instead, on traditional clinical opinion when we use these tests? For what purpose were these tests developed? If the tests give unreliable or inaccurate measurements, patients with lumbar dysfunction can be labeled inappropriately and negatively. Such labeling results in misdiagnosis, improper treatment, increased litigation, and increased cost of care. Our credibility as a profession rests on our ability to select and use reliable and valid clinical measures.

We argue that our credibility is enhanced when we use appropriate terminology to address the issue of sincerity of effort and to differentiate it from the biobehavioral aspects of delayed recovery. The term "validity," for example, often is used to address sincerity of effort.[6,7,10,11,14] Test results are described as "valid" or "invalid" based on the results of a series of tests that are alleged to test sincerity of effort. Use of the term "validity" to describe sincerity of effort is an inappropriate application of a scientific term.[22] *Validity* refers to the extent to which a measure can be used to make an inference or judgment.[23] Research establishes evidence in support of validity, which does not change with the patient's level of effort. If patients do not give a full effort during physical evaluations, then the test results represent what they were willing to do. Because patients cannot be forced to expend more effort than they are willing to expend, validity is unaffected. There is no evidence reported in the peer-reviewed literature that any of the tests designed to provide a "validity" profile of the patient are valid in the scientific sense.[22]

Other terms that may cloud our understanding of the biobehavioral factors affecting recovery include "symptom magnification" and "exaggerated pain behavior." These terms are frequently used in clinical practice, in standardized evaluations, and in the literature to identify patients who are thought to be exaggerating the severity of their medical conditions.[9,11,13-15,18,19,24] The use of this terminology is not theoretically sound. By definition, a *symptom* is a "sensation experienced by the patient."[25] There is no laboratory test or imaging technique that can measure the patient's true versus reported experience of sensation. The terms "magnification" and "exaggeration" imply that we can measure true sensations and compare these measurements with patient reports. Thus, by definition, "symptom magnification" and "exaggerated pain behavior" cannot be measured. Use of these terms, therefore, should be avoided, as they add little information that leads to improved treatment for the patient with delayed recovery.[26]

Table 1.
Widely Used Methods of Determining Sincerity of Effort

Waddell's nonorganic signs[29]
Coefficient of variation[84]
Bell-shaped curve[113]
Rapid exchange grip[106]
Correlation between musculoskeletal evaluation and functional capacity evaluation[17]
Documentation of pain behavior[54]
Documentation of symptom magnification[9]
Ratio of heart rate and pain intensity[119]

DE Lechner, PT, is Associate Professor, Division of Physical Therapy, School of Health-Related Professions, The University of Alabama at Birmingham, Bishop 102, 900 19th St S, Birmingham, AL 35284 (dlechner@uab.edu), and President and founder of ErgoScience Inc, Birmingham, Ala. Address all correspondence to Ms Lechner.

SF Bradbury, MAOM, ATC, is President and founder of Quality Essential Health Systems Inc and Co-Director of WorkHab Australia, Dallas, Tex.

LA Bradley, PhD, is Professor of Medicine, Division of Clinical Immunology and Rheumatology, The University of Alabama at Birmingham.

Table 2.
Waddell's Nonorganic Signs (NOS) Test Administration and Scoring[a]

Type of NOS	Sign	Test Administration	Positive Score
Tenderness	Superficial	Examiner pinches skin lightly between thumb and forefinger over the thoraco-lumbo-sacral region	Patient reports tenderness over a "wide area of lumbar skin"
	Deep	Examiner applies deep pressure over the thoraco-lumbo-sacral region	Patient reports tenderness over a "wide area"
Simulation tests	Axial loading	With patient standing, the examiner applies pressure to the top of the patient's head	Patient reports back pain secondary to the pressure
	Rotation	With patient standing, the examiner rotates the shoulders and pelvis as a unit	Patient reports back pain secondary to the rotation
Distraction test	Straight leg raise (flip test)	Straight leg raise is tested in a supine position, and the patient is then asked to extend the knee in a sitting position	Patient "shows marked improvement" when sitting knee extension is compared with supine straight-leg-raising test
Regional disturbances	Weakness	Manual muscle testing of the lower extremities	"Giving way" of many muscle groups that cannot be explained on a localized neurological basis
	Sensory	Sensory testing for light touch, and pinprick over the lower extremities	"Diminished sensation to light touch, pinprick . . . fitting a 'stocking' rather than dermatomal pattern"
Overreaction	Overreaction	Examiner makes observations during the examination	Disproportionate verbalization, facial expression, muscle tension and tremor, collapsing, or sweating

[a] Adapted from Waddell et al.[29]

Some patients with musculoskeletal dysfunction can give less than a full effort during physical evaluations for a variety of reasons. Pain, fear of pain, fear of reinjury, anxiety, depression, lack of understanding of instructions, lack of understanding of the importance of the test, and secondary financial gain are some of the reasons underlying self-limiting behavior. Lechner et al[27] found that therapists can reliably identify when a patient apparently is giving a full effort during functional capacity evaluations (FCEs). These investigators identified maximal effort by comparing patients' willingness to continue performing functional tasks with therapists' observations of body alignment and movement patterns. Patients were classified as participating fully, stopping before a full effort was reached, or willing to continue beyond a safe maximum effort. Sixty-two percent of the subjects self-limited their performance on 2 or more of the tasks tested, in our view, giving clinicians an adequate opportunity to judge and score self-limiting behavior. Using this approach, they achieved kappa scores for interrater reliability ranging from .56 to .97, with 95% of the scores being greater than .61.[27] According to Landis and Koch's suggested reference values, these kappa scores fall in the substantial to almost perfect agreement range.[28]

Lechner et al,[27] however, made no claim that underlying motivation was measured. If self-limiting behavior predominates during an FCE, further psychosocial or environmental evaluation may elucidate underlying factors that can be addressed during treatment, thus optimizing the likelihood of patient recovery.[26] Lack of sincerity of effort, however, should not be inferred from self-limiting participation in testing.

Waddell's Nonorganic Signs

Waddell et al,[29] in 1980, first described nonorganic signs as clinical signs that have a "predominantly nonorganic basis" in patients with LBP. They defined 8 tests for these signs and grouped them into 5 types (Tab. 2). In types with more than one test, scoring positively on one test yields a positive score for that type. According to Waddell et al,[29] a patient must score positively on 3 out of the 5 types of signs to score positively on the nonorganic signs. Waddell et al[29] stated that scoring positively on these signs identifies patients who might benefit from psychological assessment.

These signs were not intended for use in detecting sincerity of effort or malingering.[29,30] Waddell et al[29] reported that the nonorganic signs did not correlate with the F and K validity scores of the Minnesota Multiphasic Personality Inventory in 120 patients with chronic LBP from Canada and Great Britain. These investigators stated that the nonorganic signs were "not limited to, nor specific to, medicolegal and compensation situations."[29] Unfortunately, nonorganic signs are frequently used in clinical practice, in standardized evaluations, and in regulatory guidelines to imply sincerity of effort or an exaggeration of symptom reporting.[7,11,13–15] In addition, clinical investigators have strayed from the original intent of the nonorganic signs and have described their purpose as being to "detect whether patients were accurately reporting pain"[31] or to "establish the authenticity or validity of pain reports."[32]

Even when nonorganic signs are used appropriately, the premise that they are based on nonorganic symptoms is questionable. For example, the tenderness tests are scored positively if the superficial or deep tenderness does not follow a pattern associated with nerve root irritation. Regional disturbances are scored positively if the findings do not follow a dermatomal distribution.[29] Doxey et al[33] demonstrated that the nonorganic signs were associated with the absence of nerve root involvement. Such an approach seems to imply that the only valid organic source of pain is the nerve root. There is, however, considerable evidence that many structures other than the nerve root can lead to pain that extends over a broad area.[34–37] Thus, the tenderness tests may contribute to an erroneous classification of patients as needing further psychological assessment when their primary problem stems from organic sources or tissues other than nerve roots.

Several issues surrounding test administration should cause us to question the validity of nonorganic signs for detecting nonorganic problems. The simulation tests are assumed not to cause pain in the lumbar region.[29] In our clinical experience, patients with poor trunk strength or a lack of kinesthetic awareness may respond to axial loading with movement and resulting pain. The trunk rotation test presumes that the examiner is able to produce simultaneous rotation at the pelvis and shoulders without rotation occurring at the lumbar spine. In our clinical experience, eliminating motion in the lumbar region during this maneuver may not be possible, even with precise verbal instructions and careful hand placement.

The supine straight-leg-raising test and knee extension in a sitting position are presumed by Waddell et al[29] and by many clinicians to be equivalent measures. In fact, the measures are quite different. In a supine position, the spine and pelvic positions are supported by the surface on which the patient is lying, providing some stability for the lumbar spine and pelvis. In sitting, however, the spine and pelvis are unsupported.[29] The patient may slump or allow the pelvis to rotate posteriorly when the knee is extended, particularly if hamstring muscle tightness is present. Thus, in our clinical experience, knee extension in a sitting position may exceed the range of motion during the supine straight-leg-raising test in some patients who are giving a sincere effort during musculoskeletal evaluation. Although many clinicians consider only "profound" differences to be positive, the cutoff point has not been defined or examined in research studies.[29]

According to our interpretation, Waddell's overreaction test, another of the nonorganic signs, is based on the assumption that there is a standard, acceptable intensity of response to the experience of pain, against which reaction can be measured.[29] We know that past experiences, cultural background, and socioeconomic status can affect the intensity of the pain experience and expressions of pain.[38–42] Both the examiner and the patient are subject to these influences. Perhaps these factors were, at least in part, responsible for the poor reliability of this nonorganic sign.[43]

Several aspects of the nonorganic signs test administration and interpretation appear to lack clear definitions. Terms such as "wide area" in the assessment of tenderness and "disproportionate" in the overreaction test are 2 examples of terms lacking operational definitions that, in theory, could enhance the reliability and consistency of scoring the responses. Techniques of test administration, such as the amount of pressure exerted during the axial loading, hand placement during rotation, and pelvic control during the distraction test, are not standardized in Waddell's original description of the nonorganic signs[29] and may introduce variability into testing and scoring the nonorganic signs. These clinical opinions of the nonorganic signs will require further study.

Five studies addressing the reliability of scores for nonorganic signs were found (Tab. 3). Early studies conducted by Waddell and colleagues[29,44] demonstrated a high degree of interrater reliability. This reliability, however, was established among individuals who received considerable informal, nonstandardized training from Waddell. The ability to generalize the results of this study to routine clinical practice, therefore, is limited. Other investigators[43,45] have had more difficulty demonstrating the reliability of scores for nonorganic signs. The study by McCombe et al[45] involved a low proportion of positive nonorganic signs, which may have artificially lowered the kappa coefficient. Regardless of these statistical issues, these authors advise caution in interpreting superficial tenderness and abnormal sensory or motor disturbance. Future studies should report

Table 3.
Studies Examining Reliability of Noting the Presence of Nonorganic Signs (NOS)

Author (Year)	No. of Subjects	Type of Subjects	Method	Findings
Waddell et al[29] (1980)	50	Canadian WCB[a] patients	Interrater reliability between 2 physicians who independently examined patients within same week Physicians "had worked closely together for >6 mo" Test-retest reliability between admission and discharge scores (average of 23 d between evaluations)	Interrater 80% agreement Intrarater 85% agreement for identifying positive versus negative NOS
Waddell et al[44] (1982)	50	WCB patients	Patients examined independently by 2 physicians Before data collection, "detailed discussion on exact information to be gathered, and its format, qualifications, and exclusions"	Moderate to substantial reliability in assessing NOS (κ=.55–.77) Did not report statistics for reliability for each sign
Korbon et al[43,b] (1987)	39	Outpatients with chronic LBP[c]	Studied interrater reliability using 2 physicians who evaluated patients independently Attempted more quantitative version of NOS	Found axial loading (.69), rotation (.57), and overreaction (.44) to be unreliable Documenting the degree of tenderness (.48), weakness (.72), and sensory disturbance (.83) also unreliable
McCombe et al[45] (1989)	Group1: 50 Group2: 33	Outpatients with LBP	Interrater reliability between 2 orthopedic surgeons and between an orthopedic surgeon and a physical therapist	Found tenderness (κ_1=.29, κ_2=.17) and regional disturbances (κ_1=−.03, κ_2=.26) to be unreliable
Spratt et al[32] (1990)	42	LBP>4 mo	Rater pairs (one examiner, one observer) evaluated rotation, superficial tenderness, axial loading, distraction; determined reliability of judging behavioral response using standardized ordinal scoring system	Intraclass correlation coefficients for judging behavioral responses to NOS ranged from .78 to .97

[a] Workers' Compensation Board.
[b] Type of correction used not documented.
[c] LBP=low back pain.

both kappa values and percentages of agreement to address this issue. Spratt et al[32] improved the reliability of scores for nonorganic signs by using a system for coding the behavioral responses to 4 of the nonorganic signs tests: rotation, axial loading, superficial tenderness, and distraction. Their design, however, allowed 2 clinicians to observe a patient simultaneously. Such a design addresses only the reliability of interpreting responses, not the ability to obtain the measurement in clinical practice. Because much of the variability of the nonorganic signs may lie in test administration, reliability of the behavioral classification may have been overstated in this study.

The validity of nonorganic signs for predicting treatment outcomes has been examined by several investigators. Table 4 summarizes the studies that address outcome prediction. Evidence can be found that supports the validity of nonorganic signs for predicting outcomes,[29,31,33,46,47] whereas other researchers question the predictive validity of nonorganic signs.[31,33,46–53] Often, contradictory findings are reported for the same study.[31,33,46,47,50] By definition, some of the signs are scored positively if the patient's pattern of pain is diffuse rather than following a radicular pattern. Patients whose neurological symptoms improve have fewer nonorganic signs.[38,46] These findings suggest that nonorganic signs are more likely to be positive for the patient with

Table 4.
Studies Addressing the Predictive Validity of Using Waddell's Nonorganic Signs (NOS) for Prognosis

Author (Year)	No. of Subjects	Type of Subjects	Method	Findings
Waddell et al[29] (1980)	50–100[a] (170 total)	LBP[b] ranged from acute to chronic	NOS, MMPI,[c] medical history, physical examination, radiographs, disability, medicolegal and compensation factors, pain drawing	Fewer NOS in newly diagnosed patients than in patients with chronic LBP or "problem patients" Correlated with work loss (Pearson r=.30); failure of treatment (Pearson r=.19), disability, neurotic triad of MMPI (Pearson r=.18–.29) Not correlated with Workers' Compensation (equal incidence between Workers' Compensation versus non–Workers' Compensation) or medicolegal situations (Pearson r=.22), which accounted for less than 1% of variance of NOS Not correlated with validity scores of MMPI (F and K) (no correlations reported)
Lehmann et al[31] (1983)	40	Patients with chronic LBP referred to orthopedic clinic, nonsurgery candidates ≥3 mo duration, consecutive admissions	Random assignment to TENS,[d] placebo TENS, or electroacupuncture group Evaluated deep and superficial tenderness, axial loading, and rotation to distinguish "valid" (n=30) and "invalid" patients (n=10)	Invalid group correlated to gender (80% of invalid group patients were women) and having an attorney (80% of invalid group patients had an attorney) Positive NOS found to have "contamination" effect on treatment outcome (treatment×validity group interaction: F=2.04, P<.11 at discharge; F=3.66, P<.05 at follow-up) No correlation between invalid group and adherence with treatment (no data reported) No difference between valid and invalid groups on validity scales of MMPI (L, F, and K) (no data reported)
Dzioba and Doxey[46] (1984)	116	WCB[e] candidates for surgery	Evaluated with NOS (21) using same protocol as Doxey et al[33] (1988)	Fewer positive NOS correlated to relief of symptoms and neurological signs (72% of those with 0 or 1 positive NOS rated as improved), patient opinion that surgery had good result (62% of those with 0 or 1 positive NOS rated surgery as having good result), RTW (49% of those with 0 or 1 positive NOS returned to work) Combined with 4 other predictors (English proficiency, back versus leg pain, hypochondriasis of MMPI, pain drawing), 82% correct predictions of treatment outcome

[a] Series of studies.
[b] LBP=low back pain.
[c] MMPI=Minnesota Multiphasic Personality Inventory
[d] TENS=transcutaneous electrical nerve stimulation.
[e] WCB=Workers' Compensation Board
[f] RTW=return to work.
[g] NS=not significant.
[h] SLR=straight leg raise

mechanical back pain or a diffuse soft tissue pattern of pain. Other studies[29,48,49] suggest that, rather than being a predictor of outcomes, nonorganic signs may change as a result of treatment.

In summary, the nonorganic signs were not developed for the purpose of detecting sincerity of effort, and there is no evidence in the literature to suggest that they can be used for this purpose. Several researchers have raised questions regarding the reliability of the measurements. The evidence regarding the validity of nonorganic signs for predicting outcomes, such as response to treatment and return to work, is inconclusive.[31,33,46–53] In our view, nonorganic signs should be used only for the purpose for which they were originally intended; that is, to identify those persons whose physical recovery may be

Table 4.
Continued

Author (Year)	No. of Subjects	Type of Subjects	Method	Findings
Waddell et al[48] (1984)	160 main analysis cross-checked on 120 patients	LBP≥3 mo (consecutive)	Evaluated patients with pain drawings, pain scale, McGill Pain Questionnaire, multiple psychological questionnaires, physical impairment examination, severity rating of disability	NOS found to be associated with severity (explained 26.9% of variance) but explained only 5.1% of variance of chronicity and only 4.5% of variance of disability
Waddell et al[40] (1984)	380	LBP≥3 mo (consecutive)	Evaluated patients with pain drawings, multiple psychological questionnaires, NOS, history of previous treatment, extent of disability	"Illness behavior," which includes both NOS and symptoms, explained only 15% of variance of amount of treatment received
Waddell et al[49] (1986)	185	LBP	Evaluated preoperatively and postoperatively for pain, disability, and physical impairment	NOS improved in 32% of patients if surgery was successful and persisted, or worsened in 72% of patients if surgery was unsuccessful
Bradish et al[51] (1988)	120	LBP out of work 3–6 mo (consecutive)	Evaluated initially at 3–6 mo, reevaluated at 12–18 mo	Those with positive NOS had 35.1% RTW,[f] those with negative NOS had 41.0% RTW; chi-square test revealed no significant difference between these figures 27% of those with positive NOS; 32% of those with negative NOS improved; chi-square test revealed no significant difference
Doxey et al[33] (1988)	116	WCB candidates for surgery (random)	Evaluated using orthopedic and neurological examinations, NOS, MMPI, Hendler Questionnaire Pain Drawing, interview; reevaluated 12 mo later	Patients who went on to have surgery had a mean of 1.9 NOS, whereas patients without surgery had a mean of 4.4 NOS; significantly different (t=4.11, df=114, P=.0001) In patients who went on to have surgery, better pain relief and improvement in neurological deficits were inversely associated with fewer NOS (Spearman rank-order correlation=−.41, P<.05), but there was no inverse correlation to RTW (Spearman rank-order correlation=−.12, P<.05) In patients who were managed conservatively, better pain relief, improvement in neurological symptoms (Spearman rank-order correlation=−.16, P<.05), and RTW were inversely correlated with fewer NOS (Spearman rank-order correlation=−.35, P<.05)
Lacroix et al[53] (1990)	100 (sample 1: 50, sample 2: 50)	Low back injury 3–6 mo (random), WCB	Evaluated initially at 3–6 mo, reevaluated at 6–24 mo	Inconsistent correlation between NOS, RTW, and prognosis Sample 1: RTW=−.29 (NS[g]), prognosis=.17 (NS) Sample 2: RTW=−.40 (P<.05), prognosis=−.19 (NS)

Table 4.
Continued

Author (Year)	No. of Subjects	Type of Subjects	Method	Findings
Lancort and Kettelhut[50] (1992)	134	Patients with acute and chronic LBP receiving Workers' Compensation	Evaluated personal, physical, demographic, financial, and psychological factors, organic signs and NOS Discriminant analysis used to predict membership (RTW versus no RTW)	Partial r^2 of .062 for verbal magnification, .031 for superficial tenderness, and .153 for sham SLR[h] test (plantar flexion with SLR) for predicting RTW only for those off work ≤6 mo ($P<10$)
Chan et al[52] (1993)	650	LBP	Evaluated patients with pain drawings and NOS	In 127 patients with 3 or more positive NOS, 103 (81.7%) had a pain drawing that was scored as nonorganic or possibly nonorganic No significant differences were found in medicolegal status or Workers' Compensation status between patients with 0–2 NOS and patients with 3 or more positive NOS
Karas et al[47] (1997)	126	WCB patients with LBP assessed at 5 locations of Canadian Back Institute	Physical therapists administered McKenzie repetitive test movements and modified NOS, 6-mo follow-up to determine RTW	Among patients who centralized symptoms, Fisher exact test revealed more patients with low NOS scores (n=104) RTW than patients with high NOS scores (n=22) ($P=.0003$) Logistic regression revealed that for patients who centralized symptoms, probability of RTW increased with low NOS scores ($P=.0005$) Fisher exact test showed that for patients who did not centralize symptoms, NOS did not have a significant effect on RTW Patients with low NOS scores (n=104) returned to work more often than patients with high NOS scores (n=22) ($\chi^2=7.53$, $P=.006$) Patients with low NOS scores who centralized symptoms (n=79) returned to work more often than patients who did not centralize symptoms (n=25) ($\chi^2=9.29$, $P=.002$) Logistic regression: probability of RTW increased with centralization ($P=.0034$) For patients with high NOS scores, Fisher exact test showed no significant difference in rate of RTW ($P=.192$) Logistic regression revealed complex relationship between RTW and NOS and centralization of symptoms

affected by psychosocial factors and who may need psychosocial assessment or even treatment. Statements regarding sincerity of effort cannot be supported through clinical application of nonorganic signs. Caution should be exercised when using nonorganic signs to predict outcomes such as response to treatment and return to work.

Documentation of Pain Behavior

In reports of musculoskeletal evaluation or FCE of patients with LBP, clinicians frequently document that patients have "exaggerated" or "excessive" pain behavior.[6,9,11,13–15,17,18,24] As discussed in the introduction, these terms add little to our understanding of self-limiting behavior, and they negatively label patients.[26] Research indicates that an individual's pain behavior may be influenced by a variety of factors: environmen-

Table 5.
Pain Behavior Assessment Methods

Author (Date)	Pain Behaviors Assessed	Protocol	Reliability	Validity	Other Findings
Keefe and Block[55] (1982)	Bracing Guarding Rubbing Grimacing Sighing	Patients videotaped during 1- to 2-min periods of walking, sitting, standing, and reclining Raters viewed 20 20-s intervals of videotape	Interrater: percentage of agreement ranged from 93% to 99% (n=27)	Total frequency of pain behaviors correlated with pain ratings (n=27) (r=.71, P<.01)[a] Patients with chronic LBP[b] demonstrated more pain behaviors than did patients who were depressed (n=10) (t=3.65, df=35, P<.01) or subjects in comparison group (n=10) ($\chi^2[1]$=20.9, P<.01) Frequency of pain behaviors correlated[a] with naive ratings of pain (n=22) (r=.68, P<.01)	Frequency decreased with treatment
Richards et al[57] (1982)	Vocal complaints Down time Facial grimaces Standing posture Mobility Body language Support equipment Stationary movement Medication use	Patients observed without videotape (in person) while they were asked to walk a short distance, stand momentarily, and move from a sitting position to a standing position and back to a sitting position	Mean interrater reliability between nurse, psychologist, and medical student of .95 (n=50) (P<.001)[a] Test-retest reliability of .89 (P<.01) over 2 days[a] (n=50) Used 3 raters: nurse, medical student, psychologist	Correlation[a] between pain behaviors and self-report of pain intensity .16 on admission (not significant) and .55 at discharge (n=70) (P<.05) Inversely correlated[a] to daily activity level of walking, sitting, and standing (n=70) (−.30, −.29, and −.38 respectively; P<.05)	
Cinciripini and Floreen[38] (1983)	Touching Grimacing Gesturing Laughter Smiles Changing topic of conversation Fluency of speech Loudness of speech Affect	Subjects were videotaped during structured interview of 4 questions and 1 behavior task (included walking, bending, lifting, and carrying)	Spearman rank-order correlations ranged from .79 to .93 (n=25)	None reported	Environmental factors influenced observation of pain behaviors

[a] Specific correlation coefficient used not noted.
[b] LBP=low back pain.
[c] RA=rheumatoid arthritis.
[d] CHIP=Checklist for Interpersonal Pain Behavior.
[e] MMPI=Minnesota Multiphasic Personality Inventory.

tal,[38] verbal reinforcement,[39] ethnicity,[41] and interaction with one's spouse.[42] Despite these complexities, some clinicians believe that observing and reporting pain behavior has important implications for treatment and prognosis in patients with LBP. The questions clinicians must ask themselves are: (1) Do we have systematic and reliable methods of documenting pain behavior? (2) Should self-reports of pain intensity correlate with pain behavior? and (3) Is the incidence or type of pain behavior related to treatment outcomes or return to work?

An in-depth review of all the methods used to quantify pain behavior is beyond the scope of this article. The reader is referred to a recent review by Solomon[54] for a more thorough treatise on this topic. In Table 5, we briefly summarize the major approaches to quantification of pain behaviors that are pertinent to the patient with LBP. Typically, these methods rely on both verbal cues (eg, vocal complaints, moaning, sighing) and nonverbal cues (eg, bracing, guarding, rubbing, grimacing, posture, use of assistive devices) that are observed while the patient performs a structured set of tasks or during an interview. Some of the methods rely on videotaping,[38,55,56] whereas other methods are designed to be

Table 5.
Continued

Author (Date)	Pain Behaviors Assessed	Protocol	Reliability	Validity	Other Findings
Turk et al[69] (1985)	Activity avoidance Distorted ambulation Negative affect Facial/audible expression	Began with 63 pain behaviors identified through literature review Reduced list to 20 behaviors through staff consensus Used ratings from 25 psychologists and 25 physicians to cluster behaviors into the 4 categories	No reports found	No reports found	
Follick et al[56] (1985)	Similar behaviors as Keefe and Block[55] but also includes verbal statements regarding pain (16 tasks)	Subjects were videotaped during structured sequence of movements and interview	Reliability coefficients ranged from .51 to .88 over 16 pain behaviors 71 patients with chronic LBP, 40 control subjects	Four behaviors discriminate between patients and control subjects: partial movement, limitation of statements, sounds, and position shifts accounted for 75% of variance in group membership (patients versus control subjects)	Correlations between Follick's Audiovisual Taxonomy and other outcome measures (McGill Pain Questionnaire, Profile of Mood States, Beck Depression Inventory, self-handicapping [self-report of pain-related function], behavior ratings by primary nurse) ranged from .05 to .40[76,a] A hierarchical multiple regression analysis showed that pretreatment scores on Follick's Audiovisual Taxonomy did not predict posttreatment scores on other outcome measures[76]
McDaniel et al[64] (1986)	Bracing Grimacing Rubbing Sighing Rigidity Flexing extremities	Revised Keefe and Block[55] protocol for patients with RA[c]	Kappa coefficients of interrater reliability ranged from .80 to 1.00 (n=20)	Construct validity Pearson product-moment correlation coefficients between frequencies of pain behaviors and self-reports of pain and disability ranged from .26 to .45 (n=53) ($P<.001$ for 5 out of 7 measures) Pearson product-moment correlation coefficients between frequencies of pain behaviors and naive ratings ranged from .54 to .57 (n=25) ($P<.01$)	Pain behaviors decreased with cognitive therapy but not with drug treatment (n=11) (t=2.23, $P<.05$)

administered "live."[57-59] Evaluation time ranges from 5 minutes of observation[57] to 45 to 60 minutes of interviews.[59] Most of the methods have been studied for reliability,[38,55,57,58,60] with the highest levels of reliability being reported by Keefe and Block[55] and Richards et al.[57] Some of the instruments have been studied for construct validity.[55,56,58,60-64] Some methods have been found to discriminate between patients with pain and control subjects[55,56,60-64] or to be inversely correlated to functional activity level[57] and wellness behaviors.[58] Keefe and Block's method has been validated for both patients with chronic LBP[55,65-67] and patients with "nonchronic"

Table 5.
Continued

Author (Date)	Pain Behaviors Assessed	Protocol	Reliability	Validity	Other Findings
Vlaeyen et al[58] (1990)	Distorted mobility Verbal complaints Nonverbal complaints Nervousness Depression Day-sleeping	Initially assessed by nursing staff over 1-wk period of an inpatient admission Used a variety of data collection protocols, including dynamic, static, and functional testing	Pearson product-moment correlation coefficients for intrarater reliability ranged from .45 to .74 (n=46) ($P<.001$) Pearson product-moment correlation coefficients for interrater reliability ranged from .65 to .83 (n=47) ($P<.001$)	Kendall tau correlation coefficients between individual CHIP[d] scores and frequency counts from videotape, UAB Pain Behavior Scale (Richards et al,[57] 1982), and Audiovisual Taxonomy for Assessing Pain Behavior (Keefe and Block,[55] 1982) ranged from .43 to .81 (n=19) (4 out of 5 correlations significant at $P<.001$) Pearson product-moment correlation coefficients between CHIP scores and measures of health-related behavior such as sitting, standing, walking, and bicycling ranged from −.62 to .07 (n=46) (6 out of 12 correlations significant at $P=.01$ or better)	Pearson product-moment correlation coefficients between CHIP scores and the MMPI[e] depression scale and Welsh Anxiety Scale ranged from .07 to .49 (n=46) (4 out of 10 correlations significant at $P<.01$ or better)
Dirks et al[59] (1993)	Limping Grimacing Moaning Rigid torso Torso or pain site stretch Bracing when seated Weight shifting when seated Slumped posture Stands Has to walk Down time Spontaneous speech Difficulty arising Difficulty sitting Uses mechanical help Clutches or rubs pain site	Assessed during 45- to 60-min interview Yes/no scoring (9 items) Ordinal scoring (7 items)	None reported	None reported	

LBP.[68] Several instruments, developed to evaluate pain behavior, have not undergone reliability or validity testing.[38,59,69]

When pain behaviors are inconsistent with self-reports of pain, the patient often is viewed with suspicion.[54] The literature addressing the correlation between observed pain behaviors and self-reports of pain intensity, however, is contradictory. Some investigators[57,59] have found little correlation between these 2 measures, whereas other investigators[55,64] have reported a high correlation between these measures. These seemingly contradictory findings may be explained by the timing and context of the assessments, the type of behavior assessed,[70] the nature of the diagnoses, the method of assessment used, environmental characteristics,[38] and even the physical appearance of the patient.[71] When discrepancies between pain intensity and pain behaviors arise, one measure does not necessarily invalidate the other measure. Instead, we believe these discrepancies should serve as a catalyst for further evaluation that will assist in directing treatment.[70]

The relationship between overt pain behaviors and treatment outcomes also is not well established. Hasenbring et al[72] found that pain behaviors in patients with acute lumbar disk prolapse or protrusion predicted persistent pain 6 months after injury but did not predict application for early retirement. In a sample of 17 patients with chronic LBP, pain behavior during the first epidural block procedure predicted treatment outcomes.[73] The investigators suggested, however, that these results be viewed with caution due to the small sample size. Devulder et al[74] found no correlation between treatment outcomes and pain behavior in patients treated with epidural injections. Hazard et al[75] found no correlation between disability exaggeration and work status. Kleinke and Spangler[76] found no correlation between pain behaviors and other measures of outcome.

In persons with chronic LBP, pain behavior has been noted to improve with treatments that directly address behavioral issues, such as cognitive behavioral therapy (CBT).[77,78] Linton et al[77] combined a cognitive behavioral approach and exercise as a secondary preventive approach for 66 nurses who had sustained a back injury. They found that the intervention group decreased the number of pain behaviors observed during a 10-minute standardized activity session by half, whereas a waiting-list control group increased the number of pain behaviors. The intervention group showed significantly less pain behavior after treatment than did the control group (P=.007). Researchers also have demonstrated successful treatment with CBT in patients with rheumatoid arthritis[62] and in patients with fibromyalgia syndrome.[79] In contrast, pain behavior did not respond to treatment with CBT in patients with osteoarthritis of the hip and knee.[80] With other treatments, such as electrical stimulation implants[81] and drug treatment,[82] pain behaviors do not appear to change, despite improvements in self-reports of pain.

In summary, pain behavior is only one aspect of the complex experience of pain.[26] Failure to include measures that address other cognitive and psychosocial variables that affect the pain experience may result in an incomplete assessment and inappropriate treatment of patients.[26,70] Informal, nonstandardized descriptions of pain during routine clinical practice are subject to considerable error and bias.[83] A variety of pain behavior scales can be administered in a brief period, and some scales have clearly established psychometric properties.[55,57,65–68] Clinicians should choose instruments that yield measurements with the highest levels of reliability and validity, consistent with the Standards for Tests and Measurements in Physical Therapy Practice.[22] Still unknown, however, is the correlation of these measures to sincerity of effort or treatment outcomes. Further research is needed before the pain behavior scales can be used for these predictions. Care should be taken, therefore, when interpreting the results of pain behavior scales. Given our current knowledge, pain behavior scales should not be expected to correlate with reports of pain intensity or with outcomes. Their most appropriate use is for documentation of progress toward behavioral goals during rehabilitation. If patients cite pain as a reason for giving a submaximal effort on a majority of tasks of a functional assessment, a comprehensive pain assessment, which is beyond the scope of this publication, is recommended.[26]

The Use of the Coefficient of Variation in Muscle Performance Tests

The coefficient of variation (CV) is a measure of relative standard deviation, in which the standard deviation is divided by the mean and multiplied by 100 to express the standard deviation as a percentage of the mean.[84] In clinical practice, patients sometimes are asked to perform a maximal voluntary contraction repeatedly while using a strength testing device. As reported by Simonsen, "The ratio of the standard deviation divided by the mean is multiplied by 100 to yield a unitless percentage."[85(p516)] This measure frequently is used by clinicians to determine whether a patient is giving a consistent effort during testing,[2,7–9,11,16–18] which is then interpreted as a measure of sincerity of effort. The CV is not a statistic that accurately reflects reliability.[23] Instead, the CV describes the variability within a sample, some component of which is measurement error and some component of which is variability among subjects.[23] The assumption is that an intentionally submaximal effort will result in greater variability than a maximal effort.

Several problems, however, exist with the use of the CV in isometric force testing. First, there seems to be a great deal of variability in the CVs reported by different investigators and considerable discrepancy as to exactly how much variability is acceptable.[86–89] Coefficients of variation ranging from 5.1% to 29% have been reported in the peer-reviewed literature.[86–89] Using the CV to determine maximal effort requires that there be a threshold or cutoff point above which the contractions are considered submaximal. The wide variability in reported CVs makes establishing this threshold difficult. Even when submaximal efforts produce higher CVs than those produced by maximal efforts, investigators have failed to establish a definitive threshold or cutoff point, above which the effort could be classified as submaximal.[90,91]

Even if a threshold could be established, it cannot be assumed that patients who give an inconsistent effort are consciously trying to do so. High CVs can occur due to a variety of reasons, such as the type of muscle contraction used, the presence of impairment or pain, the testing protocol or equipment, or the magnitude of the variable in question.[85] Steiner et al[92] reported higher CVs for patients with painful knee syndromes than for subjects without pain when testing eccentric isokinetic knee flexion. Tornvall[93] found that the more joints and muscle groups involved in isometric testing, the higher the CVs. Coefficients of variation for lifting tasks and grip strength, therefore, may be higher than those for elbow or knee flexion. When hand-held dynamometry is used, the examiner's variability may be indistinguishable from the patient's variability.[85] In addition, the magnitude of the variable in question may influence the magnitude of the CV. Thus, individuals who are capable of producing greater force during isometric or isoinertial force testing will have a lower CV than an individual who is capable of exerting less force, merely due to the nature of the statistic itself.[23]

Based on a review of 88 English-language, peer-reviewed studies published before March 1992, Newton and Waddell concluded that "there is no evidence that [using] iso-machines [to determine CVs] provides a reliable or valid method to assess effort or to detect if the person is faking."[4(p808)] Studies cited in this review demonstrated no differences in CVs between maximal and submaximal efforts[94] and showed that subjects could reliably repeat a submaximal effort when performing isometric contractions for plantar flexion, hip flexion, and a lifting task.[95] Subsequent to this review, Newton et al[96] found measures of consistency of effort in isokinetic isoinertial testing to be unreliable and incapable of distinguishing between maximal and submaximal efforts. Simonsen[85] examined the correlation between CVs for static tasks of the ERGOS Work Simulator.* He found differences in the mean CVs between tasks and concluded that CVs will vary depending on the tasks being tested. These studies[85,94–96] and the cited review[4] provide convincing evidence that the CV cannot be used to determine sincerity of effort.

In summary, use of the CV to measure sincerity of effort is unsubstantiated in the literature. Although there is evidence that submaximal efforts can be reliably reproduced, measures of CV vary greatly depending on the instrumentation used, the task performed, the muscle groups tested, and the presence of pain. Additionally, there is little agreement between investigators and clinicians regarding the appropriate threshold CV for determining submaximal effort. Finally, there is little, if any, theoretical basis for using the CV as an index of reliability.[23] For these reasons, the CV should not be used to determine sincerity of effort.

Correlation Between Musculoskeletal Evaluation and FCE

A lack of correlation between measures of impairment determined during musculoskeletal evaluations and more functional measures is sometimes used as evidence that the patient is giving less than full effort during testing.[17] Research, however, suggests that impairment is not directly (ie, linearly) correlated with function.[97–99] Range-of-motion (ROM) measures have been found to correlate only moderately with measures of function.[100,101] Lankhorst et al[102] found that muscle force production is not correlated with function. Newton et al[96] found that isokinetic trunk muscle performance was poorly correlated with measures of disability. Roberson et al[103] found that measures of isometric and isoinertial trunk extensor performance cannot be used to predict lifting ability. MacKenzie et al[104] found low correlations between both lower-extremity ROM and muscle force and function at work and at home in patients with lower-extremity fractures. Waddell et al[105] studied patients with chronic LBP and found that a combination of measures of spinal ROM, straight leg raising, spinal tenderness, and sit-ups explained only 25% of the variance of disability.

In summary, there are no studies confirming a direct (ie, linear) relationship between musculoskeletal impairments and function. As with CVs, further research would need to establish the relationship between impairment and function as well as the thresholds or cutoff points that can be used for determining sincerity of effort.

Grip Measures

Three approaches to documenting sincerity of effort using measures of isometric grip force have evolved and are widely used: (1) calculation of the CV of repeated measures,[2,8,9,11,12,16,17] (2) analysis of force-handle position curves (bell-shaped curve),[7,11,13,17,20,21] and (3) comparison between slow sustained measures and rapid assessment measures, known as the rapid exchange grip (REG).[106] Other, less widely used methods include (1) force-time curves,[107,108] (2) rapid simultaneous grip (RSG),[109] and (3) ratio of peak and average force and ratio of slope and peak force.[108,110] Table 6 summarizes the studies that address the methods for detecting sincerity of effort with grip force.

The idea that measures of isometric grip force could be used to determine sincerity of effort was first suggested by Bechtol[111] in his initial report describing a grip dynamometer with adjustable handle spacing to accommodate a variety of hand sizes. An important aspect of

* Work Recovery Inc, 2341 S Friebus, Suite 14, Tucson, AZ 85713.

Table 6.
Studies Evaluating Grip Performance as a Measure of Sincerity of Effort

Author (Year)	No. of Subjects	Mean Age of Subjects (y)	Approach Studied	Method	Findings
Bechtol[111] (1954)	437, noninjured	Not reported	CV[b]	Described new device to measure grip strength with adjustable handle positions	First description of grip dynamometer with adjustable handle position Described factors affecting grip: heredity, size of object, exercise, handedness, time of day, day to day, age, and pain Suggested variability >20% indicates insincere effort with no supportive data
Robinson et al[2] (1993)	29, noninjured	25.2	CV	Examined detecting the sincerity of effort using CV at 50% and 100% effort	ICCs[d] for maximal and submaximal efforts, .036 and .025, respectively 55% to 69% misclassification rate Concluded CV based on 3 repetitions not adequate for determining sincerity of effort
Stokes[113] (1983)	NA[a]	Not reported	Bell-shaped curve	Described bell-shaped curve method	Description of method only
Stokes et al[114] (1995)	129, injured and noninjured	34	Bell-shaped curve REG[c]	Compared reliability of bell-shaped curve and REG Combined bell-shaped curve and REG for determining the sincerity of effort	Odd-even split correlations were .80 for static grip and 97 for REG Using a combination of bell-shaped curve and REG accurately categorized 90.6% of patients at 68% confidence level
Niebuhr and Marion[115] (1987)	25, noninjured	Junior and senior health care professional students	Bell-shaped curve	Grip strength tested in 5 handle positions Subjects asked to give full maximum effort or feign weakness	Findings did not support Stokes'[113] 1983 premise that feigned maximal effort would yield equal grip force at all 5 handle positions Curve shape differences were noted
Niebuhr and Marion[116] (1990)	30, noninjured	Undergraduate college students	Bell-shaped curve	Bell-shaped curve tested at 100%, 90%, 70%, 50%, and 30% of maximum effort	Patterns of feigned and sincere efforts were different, but findings were not significant
Smith et al[107] (1989)	43, noninjured	55	Force-time curve	Subjects asked to perform 5-s maximal effort with one hand and 50% effort with other hand	Force-time calculations involving peak and average variables correctly classified more than 92% of male subjects and 87% of female subjects

[a] NA=not applicable
[b] CV=coefficient of variation.
[c] REG=rapid exchange grip.
[d] ICC=intraclass correlation coefficient.
[e] WCB=Workers' Compensation Board.
[f] RGS=rapid simultaneous grip.

Table 6.
Continued

Author (Year)	No. of Subjects	Mean Age of Subjects (y)	Approach Studied	Method	Findings
Chengalur et al[108] (1990)	60, upper-extremity injuries	25	Force-time curve	Subjects instructed to give 100%, 75%, 50%, and 25% of maximum effort for 5 s Average and peak forces calculated from force-time curves Five discriminators examined: $D_1 = \frac{\text{Average force}}{\text{Peak force}} \times 100$ $D_2 = \frac{\text{Standard deviation of plateau}}{\text{Mean of plateau}} \times 100$ D_3 = Sincere ratio (D_1) for opposite hand minus D_1 for hand tested $D_4 = \frac{\text{Peak force} - \text{average force} \times 100}{\text{Peak force} \times \text{average force}}$ $D_5 = \frac{\sqrt[4]{\text{Peak force}} - \sqrt[4]{\text{Average force}}}{\text{Peak force} \times \text{average force}} \times 10^8$	Men: 2-factor analysis of variance (ANOVA) revealed significant difference between sincere and faking trials for D_1, D_2, D_3, and D_4 and no significant differences between trials; differences between sincere effort and 50% effort for D_1, D_2, D_3, D_4, and D_5 were 17.30, 3.56, 17.25, 0.82, and 66.86, respectively Women: 2-factor ANOVA for hand condition and level of effort revealed findings similar to those for men; differences between sincere and faking trials for D_1, D_2, D_3, D_4, and D_5 were 14.93, 3.31, 12.23, 1.25, and 124.90, respectively Sincere efforts correctly catagorized 96.7% of the time for men using D_1, D_3, D_4, and D_5; faking efforts for men correctly classified 85% of the time using D_3 and D_5 Sincere efforts for women correctly identified 100% of the time using D_1, D_4, and D_5; faking efforts for women identified 83.3% of the time using D_3 and D_5
Gilbert and Knowlton[110] (1983)	36, noninjured	Not reported	Force-time curve	Force-time curve analyzed to determine sincere and feigned efforts using contractions sustained over 5 s	For women: average force/peak force × 100 (DEV) only variable found to be significant discriminator, correctly classifying 93% of subjects (standardized and nonstandardized canonical discriminant function coefficients were 1.0 and 16.32, respectively) For men: DEV and slope to peak force were found to be best combination of discriminating variables, correctly classifying 85% of subjects (standardized and nonstandardized canonical discriminant function coefficients were 1.20275 and 9.65 for DEV and 1.00092 and 0.71 for slope)

Table 6.
Continued

Author (Year)	No. of Subjects	Mean Age of Subjects (y)	Approach Studied	Method	Findings
Hildreth et al[117] (1989)	Part I: 100, noninjured Part II: 45, patients with hand injuries Part III: 15, noninjured, instructed to feign hand injury Part IV: 45, patients with hand injuries (retrospective)	Not reported		Subjects performed alternating gripping "as fast as possible" and slow sustained gripping REG was compared with slow sustained grip	REG 15% (±1%) lower than slow sustained grip in noninjured subjects exerting maximal effort REG 79% (±67%) greater than slow sustained grip in subjects instructed to feign injury Patients with hand injuries produced bell-shaped curve REG in patients receiving Workers' Compensation was higher than in noncompensated patients Extremely high standard deviations noted
Joughin et al[109] (1993)	57, noninjured 30, patients with partial hand amputations 14 WCB* patients with hand weakness of unknown etiology	Noninjured: 38 y Patients with amputations: 37 y WCB patients: 33 y	REG and RSG[f]	Tested RSG and REG in subjects instructed to perform maximum and feigned efforts	If REG ≥25% increase over sustained grip, REG considered positive If RSG ≥16% increase over sustained grip, RSG considered positive

this study was the establishment of the relationship of force generation and handle position and the variables that affected consistency in testing grip force. Mathiowetz et al[112] later standardized the test position (subjects seated with shoulder adducted and neutrally rotated, elbow flexed to 90°, forearm in neutral, and wrist in 0°–30° of extension and 0°–15° of ulnar deviation) and procedure (3 successive trials recorded with dynamometer set to the second handle position). They also discussed reliability and validity but did not address the issue of detecting sincerity of effort.

The CV has been discussed as it relates to isometric force testing of the trunk and extremities. Further discussion of the validity of the CV in grip tests, however, seems warranted due to its wide clinical use.[2,8,9,11,12,16,17] Despite the widespread use of this approach, only Robinson et al[2] have addressed the scientific value of this protocol. In this study, both submaximal and maximal efforts demonstrated low intraclass correlation coefficients, indicating poor stability of these measures. Robinson et al also found that a high percentage of the submaximal efforts were incorrectly classified as maximal, indicating that this approach has a high percentage of false negatives. These investigators concluded that "using individual CV measures of effort consistency is not to be recommended."[2(p49)]

The bell-shaped curve approach was developed by Stokes.[113] He found that patients he considered to have a "true" hand injury showed a pattern over the dynamometer's 5 handle positions, with the greatest force exerted at the middle handle position and less force exerted at the wider and narrower positions. In contrast, he found that patients he suspected of giving an insincere effort demonstrated equal force at each of the 5 handle positions, yielding a flat, horizontal linear relationship. One problem with his study, however, was that Stokes provided no evidence of absence of disease in patients he judged as giving an insincere effort. Factors such as pain could have influenced the ability of the patients to produce more force at the middle handle position.

In a later study, Stokes at al,[114] studied subjects without pain who were asked to exert sincere and insincere efforts and patients with hand pain who were considered sincere and insincere. Both insincere groups (subjects without pain and patients) were found to have lower variability in force production at the 5 handle positions (flatter bell-shaped curve). One problem with both studies conducted by Stokes and colleagues[113,114] is that no objective process for identifying "insincere" patients was reported. Niebuhr and Marion[115,116] subsequently studied the bell-shaped curve in subjects without pain who were instructed to exert true and feigned maximal efforts, but they were unable to replicate the findings of Stokes and colleagues.[113,114]

In current clinical application, the judgment of whether the curve has an adequate bell shape and, therefore, whether a person is exerting a sincere effort or an insincere effort, typically is made by visual observation of the curve. This judgment, therefore, is nothing more than an individual clinician's opinion. The research addressing the value of using the bell-shaped curve for detecting sincerity of effort was conducted using complex statistical trend analysis techniques that few clinicians have at their disposal.

The force-time curve is a less well known approach to the detection of sincerity of effort and was first described by Smith et al.[107] Subjects are asked to sustain a maximal grip for 5 seconds. In a study of subjects without pain using the Smith protocol,[108] force-time curves for sincere-effort trials demonstrated an initial rapid rise in force that was sustained over 5 seconds. In insincere efforts, the initial rapid rise in force was followed by a "relatively gradual decline over the last few seconds."[108(p150)] Subjects who gave sincere efforts were found to have less deviation between peak force and average force compared with subjects who gave insincere efforts. A 2-factor analysis of variance revealed differences in ratios of peak and average force production between sincere and insincere efforts. The sensitivity for discriminating between sincere and insincere efforts was 90%, and the specificity for discriminating between sincere and insincere efforts was 85%. As with the bell-shaped curve, the discriminative ability of the force-time curve was not achieved using visual analysis of the data. Instead, statistical applications were used. In order for this approach to be useful and efficient for the clinicians, access to trend-analysis software would be necessary.

The REG and RSG tests are based on the premise that individuals have more difficulty maintaining a submaximal effort when the speed of grip force testing is increased from 1 sustained squeeze to 80 to 90 squeezes per minute. Those individuals who are not giving a sincere effort are supposed to show higher forces with REG testing than with slow sustained grip force testing. Consensus on the threshold for insincere effort, however, has not been achieved. Early test developers[106,117] had very vague criteria for determining a positive test versus a negative test. Czitrom and Lister,[118] for example, defined a positive REG as one in which the REG is "dramatically higher" than static grip measures. No operational definition or data were provided to quantify a specific cutoff point. By comparison, Hildreth et al[117] found a 79% greater REG in subjects who were instructed to feign injury, whereas Joughin et al[109] considered a 25% increase in REG over sustained grip force to be "positive." In addition, extremely high standard deviations for REG (±60%) were found by Hildreth et al.[117]

Joughin et al[109] modified the REG protocol such that rapid gripping of the dynamometer was performed simultaneously with both hands. The premise of this modification, termed the "rapid simultaneous grip," was that bilateral gripping made it even more difficult to differentiate the performance of each hand. These investigators determined cutoff points based on the calculated sensitivities and specificities of each test. They recommended different thresholds for patients with hand impairments and for subjects without hand impairments (Tab. 6).

The ratios of average and peak force, peak force and body weight, and slope and peak force were studied by Chengalur et al[108] and Gilbert and Knowlton.[110] They investigated the effectiveness of the ratios of slope and peak force, average and peak force, and peak force and body weight for determining sincere versus insincere efforts. Chengalur et al[108] found differences in ratios of average and peak force between subjects who gave sincere efforts and subjects who gave insincere efforts. The magnitude of the differences between sincere and "faking" trials for these ratios ranged from 0.82 to 66.86. The accuracy for predicting true sincere trials ranged from 90.0% to 96.7%, whereas the accuracy for predicting true "faking" trials ranged from 22.5% to 76.7%. Gilbert and Knowlton[110] found that for female subjects, the ratio of average and peak force was the only discriminator between subjects who gave sincere efforts and subjects who gave insincere efforts, correctly classifying 94% of the subjects. The ratio of average and peak torque and the ratio of slope and peak torque were found to correctly classify 85% of the male subjects. Their calculations of test-retest reliability yielded correlation coefficients ranging from .80 to .95 for subjects who gave insincere efforts, which suggests that these subjects can consistently reproduce submaximal efforts.

None of the studies addressing the use of grip force measures in detecting sincerity of effort were conducted with patients with LBP. The validity of measurements obtained with these methods for these patients, therefore, is unknown. Studies addressing the validity of measurements obtained with a variety of grip force measures for detecting sincerity of effort in persons with and without hand dysfunction further suggest that none of these measures have been validated adequately for this purpose.[1] Coefficients of variation calculated over 3 grip force trials are unstable and produce a high percentage of false negatives. Studies addressing the validity of the bell-shape curve approach are contradictory. The research surrounding REG is controversial regarding the cutoff scores for identifying sincere versus insincere efforts. Force-time curves, ratios of average and peak torque, and ratios of slope and peak torque are valid only when they are analyzed statistically, using techniques not currently used by most clinicians. Assessment of sincerity of effort using force-time or bell-shape curves is not valid when based on visual assessment of the data. Clinicians are advised to avoid using the CV, REG, and bell-shaped curve approaches for detecting sincerity of effort, as the literature does not support the reliability and validity of their measurements for this purpose. If force-time curves or ratios between peak force, average force, and slope are used, the results should be interpreted with caution and analyzed statistically as described by the test developers.

Relationship of Heart Rate to Pain Intensity

Another method used to determine the sincerity of effort during FCEs is to compare ratings of pain intensity on a visual analog scale (VAS) or verbal ratings with the heart rate. The premise is that as pain increases, the heart rate also increases, and when patients report high pain scores without concomitantly high heart rates, they are consciously trying to exaggerate their pain. Research conducted by Borg et al[119] is sometimes cited to justify this approach. Careful examination of Borg and colleagues' work, however, leads to questions about this interpretation. The relationship between heart rate and pain in this study was established for patients with angina pectoris exercising with a very specific bicycle ergometry protocol. The ischemic pain of angina pectoris and its associated cardiovascular consequences are somatically and physiologically different from musculoskeletal pain. The physiologic relationship among chest pain, heart rate, and workload are not the same as the relationship among LBP, heart rate, and workload. The results of Borg's studies, therefore, cannot be extrapolated to patients with musculoskeletal dysfunction.

Other studies[120–122] have addressed the heart rate response to acutely painful stimuli. The researchers found mild transient increases in heart rate in response to painful heat or cold stimuli. Coghill et al[122] found that when a painful stimulus is merely anticipated, the anxiety alone may increase heart rate. This study has important implications for correlations between pain ratings and heart rate because, in these authors' clinical experience, most of the individuals with LBP are anxious in anticipation of pain during examination.

When interpreting the results of these studies, clinicians should be aware that (1) they involved individuals without pain, (2) the painful stimulus, either heat or cold, was applied with rapid onset and rapid cessation, (3) the heart rate increases, although statistically significant, were small (5–10 beats per minute) and transient (lasting only seconds), and (4) the heart rate monitors used were more accurate and sophisticated devices than the typical chest strap models that are commonly used in physical therapy clinics. For these reasons, generalization of the results of these studies to persons with LBP is limited.

Only one group of researchers addressed the physiological response to pain in patients with chronic pain. Peters and Schmidt[123] studied physiological responses to repeated acutely painful mechanical stimuli in persons with and without chronic LBP. In response to the painful stimuli, both groups had increased skin conductance fluctuations and respiratory rates, but they did not demonstrate increases in heart rate. No correlation was found between the responses to painful stimuli and the heart and respiratory rates in either the subjects with chronic LBP or the subjects without chronic LBP. These findings suggest that heart rate cannot be used to validate self-report measures of pain.

In summary, using autonomically mediated physiological measures to validate self-report measures of pain is attractive to therapists who seek objective measures of pain. Studies supporting this relationship in individuals without LBP receiving an acutely painful thermal stimulus cannot be generalized to individuals with LBP. A relationship between a mechanical painful stimulus and heart rate has not been found in patients with chronic LBP. Preliminary research suggests a complex and poorly understood relationship among pain intensity, physiologic responses, and pain perception.[120–123] Based on these studies, accusing patients of exaggerating their pain response due to a lack of concomitant rise in heart rate is not appropriate.

Clinical Implications

The very concept or construct of sincerity of effort is illusive and difficult to measure. The definition explicitly implies a measurement of motivation. To date, none of the previously discussed methods for detecting sincerity of effort have been adequately studied for its use with patients with LBP. The medicolegal and ethical implications of this lack of validation are tremendous. Clinical reports that imply the patient has intentionally given less than a full effort are in clear violation of American Physical Therapy Association (APTA) measurement standards[22] and often have extremely negative consequences for patients with LBP. Patients may have Workers' Compensation payments withheld, may lose their jobs, may receive a diminished medicolegal settlement, or may undergo inappropriate clinical treatment as a result of the negative labeling associated with the approaches discussed. Any methods used to make determinations regarding sincerity of effort, therefore, should have research published in the peer-reviewed literature

95 Carlsoo S. With what degree of precision can voluntary static muscle force be repeated? *Scand J Rehabil Med.* 1986;18:1–3.

96 Newton M, Thow M, Somerville D, et al. Trunk strength testing with iso-machines, part 2: experimental evaluation of the Cybex II back testing system in normal subjects and patients with chronic low back pain. *Spine.* 1993;18:812–824.

97 Yelin EH, Henke CJ, Epstein WV. Work disability among persons with musculoskeletal conditions. *Arthritis Rheum.* 1986;29:1322–1333.

98 Althouse HL. Preplacement screening: revealing a true profile of musculoskeletal abilities. *Occup Health Saf.* 1980;49(5):25–30.

99 Jette AM. Concepts of health and methodological issues in functional status assessment. In: Granger CV, Gresham GE, eds. *Functional Assessment in Rehabilitation.* Baltimore, Md: Williams & Wilkins; 1984:46–64.

100 Bergstrom G, Aniansson A, Bejelle A, et al. Functional consequences of joint impairment at age 79. *Scand J Rehabil Med.* 1985;17:183–190.

101 Badley EM, Wagstaff S, Wood PH. Measures of functional ability (disability) in arthritis in relation to impairment of range of joint movement. *Ann Rheum Dis.* 1984;43:563–569.

102 Lankhorst GJ, Van de Stadt RJ, Van der Korst JK. The relationships of functional capacity, pain, and isometric and isokinetic torque in osteoarthritis of the knee. *Scand J Rehabil Med.* 1985;17:167–172.

103 Roberson JM, Witt P, Gross MT. A comparison of trunk extensor strength and squat lifting ability. *J Orthop Sports Phys Ther.* 1997;25:137–144.

104 MacKenzie EJ, Cushing BM, Jurkovich GJ, et al. Physical impairment and functional outcomes six months after severe lower extremity fractures. *J Trauma.* 1993;34:528–538.

105 Waddell G, Somerville D, Henderson I, Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine.* 1992;17:617–628.

106 Lister G. *The Hand: Diagnosis and Indications.* 3rd ed. New York, NY: Churchill Livingstone Inc; 1993:162–163.

107 Smith GA, Nelson RC, Sadoff SJ, Sadoff AM. Assessing sincerity of effort in maximal grip strength tests. *Am J Phys Med Rehabil.* 1989;68:73–80.

108 Chengalur SN, Smith GA, Nelson RC, Sadoff AM. Assessing sincerity of effort in maximal grip strength tests. *Am J Phys Med Rehabil.* 1990;69:148–153.

109 Joughin K, Gulati P, Mackinnon SE, et al. An evaluation of rapid exchange and simultaneous grip tests. *J Hand Surg [Am].* 1993;18:245–252.

110 Gilbert JC, Knowlton RG. Simple method to determine sincerity of effort during a maximal isometric test of grip strength. *Am J Phys Med.* 1983;62:135–144.

111 Bechtol C. Grip test the use of a dynamometer with adjustable handle spacings. *J Bone Joint Surg Am.* 1954;36:820–832.

112 Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. *J Hand Surg [Am].* 1984;9:222–226.

113 Stokes HM. The seriously uninjured hand: weakness of grip. *J Occup Med.* 1983;25:683–684.

114 Stokes HM, Landrieu KW, Domangue B, Kunen S. Identification of low-effort patients through dynamometry. *J Hand Surg [Am].* 1995;20:1047–1056.

115 Niebuhr BR, Marion R. Detecting sincerity of effort when measuring grip strength. *Am J Phys Med Rehabil.* 1987;66:16–24.

116 Niebuhr BR, Marion R. Voluntary control of submaximal grip strength. *Am J Phys Med Rehabil.* 1990;69:96–101.

117 Hildreth DH, Breidenbach WC, Lister GD, Hodges AD. Detection of submaximal effort by use of the rapid exchange grip. *J Hand Surg [Am].* 1989;14:742–745.

118 Czitrom AA, Lister GD. Measurement of grip strength in the diagnosis of wrist pain. *J Hand Surg [Am].* 1988;13:16–19.

119 Borg G, Holmgren A, Lindblad I. Quantitative evaluation of chest pain. *Acta Med Scand Suppl.* 1981;644:43–45.

120 Moltner A, Holzo R, Strian F. Heart rate changes as an autonomic component of the pain response. *Pain.* 1990;43:81–89.

121 Kregel KC, Seals DR, Callister R. Sympathetic nervous system activity during skin cooling in humans: relationship to stimulus intensity and pain sensation. *J Physiol (Lond).* 1992;454:359–371.

122 Coghill RC, Talbot JD, Evans AC, et al. Distributed processing of pain and vibration by the human brain. *J Neurosci.* 1994;14:4095–4108.

123 Peters ML, Schmidt AJM. Psychophysiological responses to repeated acute pain stimulation in chronic low back pain patients. *J Psychosom Res.* 1991;35:59–74.

124 Hart DL, Isernhagen SJ, Matheson LN. Witness: the truth please. *Work.* 1997;9:295–297.

55 Keefe FJ, Block AR. Development of an observation method for assessing pain behavior in chronic low back pain patients. *Behavioral Therapy.* 1982;13:363–375.

56 Follick M, Ahern D, Aberger E. Development of an audiovisual taxonomy of pain behaviour: reliability and discriminant validity. *Health Psychol.* 1985;4:555–568.

57 Richards JS, Nepomuceno C, Riles M, Suer Z. Assessing pain behavior: the UAB Pain Behavior Scale. *Pain.* 1982;14:393–398.

58 Vlaeyen JWS, Pernot DFM, Kole-Snijders AMJ, et al. Assessment of the components of observed chronic pain behavior: the Checklist for Interpersonal Pain Behavior (CHIP). *Pain.* 1990;43:337–347.

59 Dirks JF, Wunder J, Kinsman R, et al. A pain rating scale and a pain behavior checklist for clinical use: development, norms, and the consistency score. *Psychother Psychosom.* 1993;59:41–49.

60 Anderson KO, Bradley LA, McDaniel LK, et al. The assessment of pain in rheumatoid arthritis: disease differentiation and temporal stability of a behavioral observation method. *J Rheumatol.* 1987;14: 700–704.

61 Anderson KO, Bradley LA, McDaniel LK, et al. The assessment of pain in rheumatoid arthritis: validity of a behavioral observation method. *Arthritis Rheum.* 1987;30:36–43.

62 Bradley LA, Young LD, Anderson KO, et al. Effects of psychological therapy on pain behavior of rheumatoid arthritis patients: treatment outcome and six-month follow-up. *Arthritis Rheum.* 1987;30:1105–1114.

63 Jaworski TM, Bradley LA, Heck LW, et al. Development of an observation method for assessing pain behaviors in children with juvenile rheumatoid arthritis. *Arthritis Rheum.* 1995;38:1142–1151.

64 McDaniel LK, Anderson KO, Bradley LA, et al Development of an observation method for assessing pain behavior in rheumatoid arthritis patients. *Pain.* 1986;24:165–184.

65 Keefe FJ, Wilkins, RH, Cook WA. Direct observation of pain behavior in low back pain patients during physical examination. *Pain.* 1984;20:59–68.

66 Keefe FJ, Hill RW An objective approach to quantifying pain behavior and gait patterns in low back pain patients. *Pain.* 1985;21: 153–161.

67 Keefe FJ, Crisson JE, Maltbie A, et al. Illness behavior as a predictor of pain and overt behavior patterns in chronic low back pain patients. *J Psychosom Res.* 1986;30:543–551.

68 Jensen IB, Bradley LA, Linton SJ. Validation of an observation method of pain assessment in non-chronic back pain. *Pain.* 1989;39: 267–274.

69 Turk DC, Wack JT, Kerns RD. An empirical examination of the "pain-behavior" construct. *J Behav Med.* 1985;8:119–130.

70 Turk DC, Flor H. Pain greater than pain behaviors: the utility and limitations of the pain behavior construct. *Pain.* 1987;31:277–295.

71 Hadjistavropoulos HD, Ross M, von Baeyer CL. Are physicians' ratings of pain affected by patients' physical attractiveness? *Soc Sci Med.* 1990;31:69–72.

72 Hasenbring M, Marienfeld G, Kuhlendahl D, Soyka D. Risk factors of chronicity in lumbar disc patients: a prospective investigation of biologic, psychologic, and social predictors of therapy outcome. *Spine.* 1994;19:2759–2765.

73 Connolly GH, Sanders SH. Predicting low back pain patients' response to lumbar sympathetic nerve blocks and interdisciplinary rehabilitation: the role of pretreatment overt pain behavior and cognitive coping strategies. *Pain.* 1991;44:139–146.

74 Devulder J, Bogaert L, Castille F, et al. Relevance of epidurography and epidural adhesiolysis in chronic failed back surgery patients. *Clin J Pain.* 1995;11:147–150.

75 Hazard RG, Bendix A, Fenwick JW. Disability exaggeration as a predictor of functional restoration outcomes for patients with chronic low-back pain. *Spine.* 1991;16:1062–1067.

76 Kleinke C, Spangler A. Psychometric analysis of the audiovisual taxonomy for assessing pain behaviour in chronic low back pain patients. *J Behav Med.* 1988;11:83–95.

77 Linton SJ, Bradley LA, Jensen I, et al. The secondary prevention of low back pain: a controlled study with follow-up. *Pain.* 1989;36: 197–207.

78 Turner JA. Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. *J Consult Clin Psychol.* 1982;50:757–765.

79 White KP, Nielson WR. Cognitive behavioral treatment of fibromyalgia syndrome: a follow-up assessment. *J Rheumatol.* 1995;22:717–721.

80 Keefe FJ, Caldwell DS, Williams DA, et al. Pain coping skills training in the management of osteoarthritic knee pain: a comparative study. *Behavioral Therapy.* 1990;21:49–62.

81 Strege DW, Cooney WP, Wood MB, et al. Chronic peripheral nerve pain treated with direct electrical nerve stimulation. *J Hand Surg [Am].* 1994;19:931–939.

82 Anderson KO, Bradley LA, Turner RA, et al. Pain behavior of rheumatoid arthritis patients enrolled in experimental drug trials. *Arthritis Care Research.* 1994;7(2):64–68.

83 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.

84 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* East Norwalk, Conn: Appleton & Lange; 1993:680.

85 Simonsen JC. Coefficient of variation as a measure of subject effort. *Arch Phys Med Rehabil.* 1995;76:516–520.

86 Agre JC, Magness JL, Hull SZ, et al. Strength testing with a portable dynamometer: reliability for upper and lower extremities. *Arch Phys Med Rehabil.* 1987;68:454–458.

87 Chaffin DB, Herrin GD, Keyserling WM. Preemployment strength testing: an updated position. *J Occup Med.* 1978;20:403–408.

88 Montgomery LC, Douglass LW, Deuster PA. Reliability of an isokinetic test of muscle strength and endurance. *J Orthop Sports Phys Ther.* 1989;10:315–322.

89 Kannus P. Normality, variability, and predictability of work, power, and torque acceleration energy with respect to peak torque in isokinetic muscle testing. *Int J Sports Med.* 1992;13:249–256.

90 Bohannon RW. Differentiation of maximal from submaximal static elbow flexor efforts by measurement variability. *Am J Phys Med.* 1987;66:213–218.

91 Saunders N, Bohannon R. Can feigned maximal efforts be distinguished from maximal efforts? *Journal of Human Muscle Performance.* 1991;1:16–24.

92 Steiner LA, Harris BA, Krebs DE. Reliability of eccentric isokinetic knee flexion and extension measurements. *Arch Phys Med Rehabil.* 1993;74:1327–1335.

93 Tornvall G. Assessment of physical capabilities. *Acta Physiol Scand Suppl.* 1963;201:1–102.

94 Robinson M, Macmillan M, O'Connor P, et al. Reproducibility of maximal versus submaximal efforts in an isometric lumbar extension task. *J Spinal Disord.* 1991;4:444–448.

12 Jacobs J. *BTE Clinical Applications Manual: A Clinically Oriented Reference Manual for Users of the BTE Work Simulator.* Baltimore, Md: BTE Co; 1992:3-1–3-12.

13 Saunders RL. *Industrial Rehabilitation: Techniques for Success.* Chaska, Minn: The Saunders Group; 1995:34–38.

14 Saunders HD, Saunders RL. *Evaluation, Treatment, and Prevention of Musculoskeletal Disorders, Volume 1: Spine.* Chaska, Minn: The Saunders Group; 1993:44, 96, 358.

15 Saunders RL, Beissner KL, McManis BG. Estimates of weight that subjects can lift frequently in functional capacity evaluations. *Phys Ther.* 1997;77:1717–1728.

16 *ERGOS Work Simulator Administration Manual.* Tucson, Ariz: Work Recovery Inc; 1987: Interims 8.

17 Isernhagen SJ. *Isernhagen Work Systems Functional Capacity Evaluation.* Duluth, Minn: Isernhagen Work Systems; 1996:52a–52b, 83.

18 Tatom AJ. Dynametric testing of the trunk. In: Basmajian JV, Nyberg R, eds. *Rational Manual Therapies.* Baltimore, Md: Williams & Wilkins; 1993:173.

19 Ogden-Niemeyer L. The issue of abnormal illness behavior in work hardening. In: Ogden-Niemeyer L, Jacobs K, eds. *Work Hardening: State of the Art.* Thorofare, NJ: Slack Inc; 1989:83–109.

20 Magee DJ. *Orthopedic Physical Assessment.* Philadelphia, Pa: WB Saunders Co; 1997:297.

21 Bear-Lehman J. Functional evaluation of the wrist and hand. In: Nordin M, Andersson GBJ, Pope MH, eds. *Musculoskeletal Disorders in the Workplace.* St Louis, Mo: Mosby; 1997:444, 452.

22 Task Force on Standards for Measurement in Physical Therapy. Standards for tests and measurements in physical therapy practice. *Phys Ther.* 1991;71:589–622.

23 Rothstein JM, Echternach JL. *Primer on Measurement: An Introductory Guide to Measurement Issues.* Alexandria, Va: American Physical Therapy Association; 1993.

24 Monsein M, Clift RB. Pain and return to work: turning the corner. In: Isernhagen SJ, ed. *The Comprehensive Guide to Work Injury Management.* Gaithersburg, Md: Aspen Publishers Inc; 1995:546.

25 Dirckx JH, ed. *Stedman's Concise Medical Dictionary for the Health Professions.* 3rd ed. Baltimore, Md: Williams & Wilkins; 1997:845.

26 Feuerstein M, Beattie PF. Biobehavioral factors affecting pain and disability in low back pain: mechanisms and assessment. *Phys Ther.* 1995;75:267–280.

27 Lechner DE, Jackson JR, Roth DL, Straaton KV. Reliability and validity of a newly developed test of physical work performance. *J Occup Med.* 1994;36:997–1004.

28 Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.

29 Waddell G, McCulloch JA, Kummel E, Venner RM. Nonorganic physical signs in low-back pain. *Spine.* 1980;5:117–125.

30 Scalzitti DA. Screening for psychological factors in patients with low back problems: Waddell's nonorganic signs. *Phys Ther.* 1997;77: 306–312.

31 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: 625–634.

32 Spratt KF, Lehmann TR, Weinstein JN, Sayre HA. A new approach to the low-back physical examination: behavioral assessment of mechanical signs. *Spine.* 1990;15:96–102.

33 Doxey NC, Mitson GL, Dzioba RB, Lacroix JM. Predictors of outcome in back surgery candidates. *J Clin Psychol.* 1988;44:611–621.

34 Mooney V, Robertson J. The facet syndrome. *Clin Orthop.* 1976;115: 149–156.

35 Badgley CE. The articular facets in relation to low-back pain and sciatic radiation. *J Bone Joint Surg.* 1941;23:481–496.

36 Hirsch C, Ingelmark BE, Miller M. The anatomical basis for low back pain: studies on the presence of sensory nerve endings in ligamentous, capsular, and intervertebral disc structures in the human lumbar spine. *Acta Orthop Scand.* 1963;33:1–17.

37 Inman VT, Saunders JB. Referred pain from skeletal structures. *J Nerv Ment Dis.* 1944;99:660–667.

38 Cinciripini PM, Floreen A. An assessment of chronic pain behaviour in a structured interview. *J Psychosom Res.* 1983;27:117–123.

39 Cairnes D, Pasina J. Comparison of verbal reinforcement and feedback in the operant treatment of disability due to chronic low back pain. *Behavioral Therapy.* 1977;8:621–630.

40 Waddell G, Birchner M, Finlayson D, Main CJ. Symptoms and signs: Physical illness or illness behavior? *BMJ.* 1984;289:739–741.

41 Tursky B, Sternbach RA. Further physiological correlates of ethnic differences in responses to shock. *Psychophysiology.* 1967;4:67–74.

42 Block AR, Kremer EF, Gaylor M. Behavioral treatment of chronic pain: the spouse as a discriminative cue for pain behavior. *Pain.* 1980;9:243–252.

43 Korbon GA, DeGood DE, Schroeder ME, et al. The development of a somatic amplification rating scale for low-back pain. *Spine.* 1987;12: 787–795.

44 Waddell G, Main CJ, Morris EW, et al. Normality and reliability in the clinical assessment of backache. *Br Med J Clin Res.* 1982;284: 1519–1523.

45 McCombe PF, Fairbank JCT, Cockersole BC, Pynsent PB. Reproducibility of physical signs in low-back pain. *Spine.* 1989;14:908–918.

46 Dzioba RB, Doxey NC. A prospective investigation into the orthopaedic and psychologic predictors of outcome of first lumbar surgery following industrial injury. *Spine.* 1984;9:614–623.

47 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: 354–360.

48 Waddell G, Main CJ, Morris EW, et al. Chronic low-back pain, psychologic distress, and illness behavior. *Spine.* 1984;9:209–213.

49 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;11:712–719.

50 Lancort J, Kettelhut M. Predicting return to work for lower back pain patients receiving workers' compensation. *Spine.* 1992;17: 629–640.

51 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;13:557–560.

52 Chan CW, Goldman S, Ilstrup DM, et al. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. *Spine.* 1993;18:1717–1722.

53 Lacroix JM, Powell J, Lloyd GJ, et al. Low-back pain: factors of value in predicting outcome. *Spine.* 1990;15:495–499.

54 Solomon PE. Measurement of pain behaviour. *Physiotherapy Canada.* 1996;48:52–58.

to support the reliability and validity of their measurements for this purpose.

The 1993 US Supreme Court ruling of *Daubert v Merill-Dow Pharmaceuticals* addressed the legal implications of judgments of sincerity of effort. This ruling established the criteria for admissibility of evidence in courts of law. According to this ruling, expert witness testimony must be based on measurements that have established reliability and validity published in peer-reviewed journals.[124] Given the available literature, the reviewed approaches for documenting sincerity of effort do not meet the criteria established by this ruling,[124] nor do they comply with APTA measurement standards.[22]

No approach for detecting sincerity of effort has been directly correlated to outcome. If patients demonstrate a "high" CV, if they display a large number of pain behaviors, if they score positively on the nonorganic signs test, if their impairment measures do not correlate with function, or if their heart rate does not increase with pain reports, these findings do not necessarily predict poor adherence or a failure to return to work. Unfortunately, any of these statements can negatively bias health care professionals against the patient. Those patients who are negatively labeled may receive inadequate or inappropriate treatment.

The question remains: Do we need to measure sincerity of effort? If we ignore this issue, then some people may receive treatment they do not need, may fail to return to appropriate work, or may receive undeserved disability payments, all of which are costly to industry and society. If we make judgments with the currently available methods, we are taking a great risk of incorrectly classifying some patients as insincere. This misclassification may cost them their job or their medical or Workers' Compensation insurance and may negatively affect their self-esteem. As a profession, we are under pressure from referral sources to assess sincerity of effort. We seriously question whether detecting sincerity of effort is an appropriate role for clinicians.

The APTA standards for measurement and practice[22] require that clinical measurements used to detect sincerity of effort have established validity. Currently, clinicians do not have legitimate tools or methods with which to make these assessments. Any statements regarding sincerity of effort, therefore, are strictly clinical opinion. Therapists are advised to avoid the use of the discussed methods for the purpose of supporting claims of detecting sincerity or level of cooperation with evaluation. Therapists who draw unwarranted conclusions from test results are violating the rights of the person being tested.[22] Therapists also are advised to avoid reporting test results as "valid" or "invalid" based on perceived levels of cooperation and to avoid using the terms "symptom magnification" and "exaggerated pain behavior" to describe patient behavior.

We suggest, instead, that clinicians and referral sources seek alternative methods to address delayed recovery and to understand the biobehavioral factors affecting pain and disability (ie, disease conviction, perceptual-cognitive bias, perceived control, perceived disability, fear of pain, perceptions of work and family, and self-efficacy).[26] Feuerstein and Beattie[26] provided an excellent theoretical framework that can enhance our understanding of these factors. In addition, these authors reviewed some commonly used approaches for assessment of biobehavioral factors and provided examples of ways in which physical therapists can use this information in clinical practice. Further research is needed to identify the most appropriate measures and treatments and to understand their implications for clinical practice. Identifying and addressing the biobehavioral factors that affect delayed recovery, however, may provide a more proactive approach to achieving functional restoration.[26]

References

1 King PM. Analysis of approaches to detection of sincerity of effort through grip strength measurement. *Work.* 1998;10:9–13.

2 Robinson ME, Geisser ME, Hanson CS, O'Conner PD. Detecting submaximal efforts in grip strength testing with the coefficient of variation. *Journal of Occupational Rehabilitation.* 1993;3:45–50.

3 Baker JC. Burden of proof in detection of submaximal effort. *Work.* 1998;10:63–70.

4 Newton M, Waddell G. Trunk testing with isomachines, part 1: review of a decade of scientific evidence. *Spine.* 1993;18:801–811.

5 Alpert J. *State of California Functional Capacity Evaluation Guidelines for Physical Therapists.* Sacramento, Calif: California Chapter, American Physical Therapy Association; 1996:5, 6, 8.

6 Bottomly S. *Interpretive Manual to Accompany the Residual Functional Evaluation for Unscheduled Disability.* Salem, Ore: Work Hardening Association of Oregon in conjunction with the Workers' Compensation Division, State of Oregon Department of Insurance and Finance; 1989:12–16.

7 *Oregon Administrative Rules for Department of Consumer and Business Services, Workers' Compensation Division.* Salem, Ore: 1997 Chapter 436, Div 010-32.

8 Matheson LN. Functional capacity evaluation. In: Demeter SL, Andersson GBJ, Smith GM, eds. *Disability Evaluation.* St Louis, Mo: Mosby; 1996:180.

9 Matheson LN. Use of the BTE Work Simulator to screen for symptom magnification syndrome. *Industrial Rehabilitation Quarterly.* 1989; 2(2):5–31.

10 Key GL. Work conditioning and work hardening. In: Key GL, ed. *Industrial Therapy.* St Louis, Mo: Mosby; 1995:254–294.

11 Blankenship K. *The Blankenship System Functional Capacity Evaluation: The Procedure Manual.* Macon, Ga: The Blankenship Corp; 1994: 4.01–4.83, 5.01–5.17, 9.01–9.37.

73 US Department of Labor, Employment and Training Administration. *The Revised Handbook for Analyzing Jobs.* Indianapolis, Ind: JIST Works; 1991.

74 Abdel-Moty E, Fishbain D, Khalil T, et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain.* 1993;9:168–173.

75 Legg J, Pateman CM. A physiological study of the repetitive lifting capability in healthy young males. *Ergonomics.* 1984;27:259–272.

76 Legg J, Myles WS. Maximum acceptable repetitive lifting workloads using psychophysical and subjective rating methods. *Ergonomics.* 1981; 24:907–916.

77 Mital A. Comprehensive maximum acceptable weight of lift database for regular 8-hour work shift. *Ergonomics.* 1984;27:1127–1138.

78 Matheson LN. "How do you know that he tried his best?" The reliability crisis in industrial rehabilitation. *Industrial Rehabilitation Quarterly.* 1988;1:10–12.

79 Astrand I, Rodahl K. *Textbook of Work Physiology.* New York, NY: McGraw-Hill; 1970.

80 Ogden-Niemeyer L. *Procedure Guidelines for the WEST Standard Evaluation: Assessment of Range of Motion Under Load.* Long Beach, Calif: Work Evaluation Systems Technology; 1989.

81 Key GL. *Industrial Therapy.* St Louis, Mo: CV Mosby Co; 1997.