

Simultaneous Bilateral Testing: Validation of a New Protocol to Detect Insincere Effort During Grip and Pinch Strength Testing

D. Schapmire, MS

*Industrial Rehabilitation Consultants
Hopedale, Illinois*

J. D. St. James, PhD

*Department of Behavioral Sciences
Millikin University
Decatur, Illinois*

Rob Townsend, MS

*Department of Exercise Science
Millikin University*

Tom Stewart, PT

*Allegheny and Chesapeake Physical Therapists
Ebensburg, Pennsylvania*

Steven Delheimer, MD

*Neurosurgeon
Peru, Illinois*

Dan Focht, OT

*Tri-State Occupational Health
Dubuque, Iowa*

Industrial injuries commonly affect the upper extremity. When these injuries are compounded by psychosocial factors, including compensation issues, the need for valid, reproducible and objective tests is paramount. Discogenic disease, including herniated discs, usually has multiple objective findings such as reflex changes, abnormal MRI findings, or EMG findings that can be cross-correlated with distal upper extremity weakness. With cumulative trauma disorders, carpal tunnel syndrome, cubital tunnel syndrome, and shoulder and cervical disorders, however, there is a paucity of objective tests for function. Furthermore, the subjective complaints associated with these disorders are widely disseminated in lay publications. Therefore, objective testing to eliminate feigned weakness, whatever the underlying motivation, is crucial.

Some of the authors have a financial interest in one or more products mentioned in this article.

Correspondence and reprint requests to Darrell Schapmire, MS, P.O. Box 171, Hopedale, IL 61747; e-mail: <ds@x-rt.com>.

ABSTRACT: The detection of feigned weakness in hand grip strength assessment is difficult. The authors review several proposed methods and their weaknesses. A comparison of unilateral testing and simultaneous bilateral testing with the Jamar dynamometer and the Baseline pinch gauge is demonstrated as a solution. An experiment involved 100 asymptomatic subjects who were tested twice, once under instructions to give a full effort and once under instructions to feign weakness. Seven statistical criteria of noncompliance were chosen. Defining noncompliance as failing two or more of the seven criteria, 99% of the instructed noncompliant subjects were correctly classified as noncompliant. No subjects were incorrectly classified as noncompliant during instructed compliant testing. Twelve subjects failed on a single criterion. On retesting, all but one were correctly classified. One subject in the instructed noncompliant group passed all criteria. Accuracy was 99.5%, including retesting of the 12 "gray-zone" subjects.

J HAND THER. 2002;15:242-250.

DETECTION OF FEIGNED WEAKNESS

The detection of feigned weakness in grip strength testing is notoriously difficult. We review the methods that have been suggested, which have generally failed. We then suggest an approach based on simultaneous bilateral grip strength tests and report a successful experimental validation of this approach.

Coefficient of Variation

The coefficient of variation (CV) is the ratio of the standard deviation of a set of measures to their mean, times 100. In strength testing, the logic is that a high CV indicates an insincere effort. A CV of 16%, for example, would result from three efforts of 40, 50, and 60 kg (mean, 50; SD, 8.16; CV: $(8.16/50 \times 1016 = 32\%)$). Using this logic, someone who could produce a force of 60 kg on one effort, but only 40 kg on another is likely to be feigning weakness.

A severe limitation in use of the CV is that, while a large CV likely indicates an invalid effort, a small COV does not imply a valid one. Robinson et al.,¹ in a study of lumbar extension, noted that "the hypothesis that submaximal efforts cannot be reproduced as consistently as maximal [isometric] efforts was not supported." Similar conclusions in grip testing have been reported by Ashford et al.,² Dvir,³ Fairfax et al.,⁴ Niebuhr and Marion,⁵ Simonsen,⁶ and Tredgett et al.⁷ Schechtman⁸⁻¹⁰ offers recent reviews in this journal, also noting that the CV may be inflated in some injured patients.

The "Bell-shaped Curve"

A second test proposed to detect feigned weakness relies on the fact that, in a sincere effort, the grip strength is greatest at position 2 or 3 of the Jamar hand dynamometer and distinctly lower at positions 1 and 5, because of the biomechanical difficulty of producing force at the widest and narrowest positions. If the function relating force to position is flat, or U-shaped, the effort is presumed to involve feigned weakness. Stokes¹¹ stated that, using this concept, "the physician can objectively document real, as opposed to fictitious, loss of grip, and can avoid subjective or argumentative statements relative to the patient's lack of cooperation." Stokes' study was based on results obtained from an unspecified number of patients "thought to be voluntarily applying minimal grip."

Niebuhr and Marion¹² found that curve patterns did differ between trials of sincere effort and feigned weakness, "but not as strongly as expected." Lechner et al.¹³ noted that curve analysis "is made by visual observation of the curve" and is "nothing more than an individual clinician's opinion." Hamilton et al.¹⁴ found that weak subjects give flatter functions across Jamar positions. We are unaware of any numeric criteria for determining when a curve is "too flat," nor have we found any statistical descriptions of variations from the expected curve.

Logically, as with the CV, a flat grip by position curve is arguably evidence of feigned weakness, but a finding of greatest strength at positions 2 or 3 does not indicate compliance. We offer additional data on this point below.

Rapid Exchange Grip Testing

Rapid exchange grip (REG) testing was proposed by Lister as a means of detecting feigned weakness (Hildreth et al.¹⁵). If a subject is asked to produce an explosive force repeatedly, while alternating grips rapidly between the hands, then an insincere effort would be indicated if the person produced a force equal to the highest force produced during phasic testing, because an explosive force produced over a brief (1 second)

period should be less than the maximal force produced during a longer (2-3 seconds) effort. Hildreth et al.¹⁵ showed that this did occur in persons deliberately feigning weakness. Joughin et al.¹⁶ reported a sensitivity of 86% and a specificity of 97%, using a 25% increase in grip strength in REG. Rapid simultaneous grip (RSG), with simultaneous bilateral testing, were found to have slightly lower sensitivity and specificity.

Taylor and Shechtman^{17,18} attempted to standardize the administration of the REG but still reported that the test "may not be sensitive or specific enough to effectively detect sincerity of effort."

In our experience, it is difficult to control the rate of a subject's performance and hence to standardize the test. In addition, we have noted that the physical impact of the subject's hand against the Jamar dynamometer during REG testing often causes false readings of force production, even when no gripping force has been applied by the subject.

Other Methodologies

The use of force-time curves has been the most successful at detecting feigned weakness. Chengalur et al.¹⁹ compared force-time curves of compliant and deliberately noncompliant efforts and reported sensitivity rates of 85.0% and 76.6% for male and female subjects, respectively, and specificity rates of 93.3% and 90.0%. The findings of Smith et al.²⁰ were similar to those of Chengalur et al.¹⁹ Gilbert and Knowlton²¹ investigated force-time curves and reported a method that was 87.5% accurate for female and 80% for male subjects, although they did not report sensitivity and specificity of their method. Mitterhauser et al.²² investigated the amounts of force produced individually by the digits in the radial and ulnar distributions during grip testing. They report a specificity of 0.98, using a 15% cutoff for the CV.

Shechtman¹⁰ calculated the sensitivity from the reported data for this study at 0.74. Statistically significant differences were found in "total pattern" with regard to force produced. While these results are certainly promising, they rely on dynamometers interfaced with a computer, and there seems to be no commercially available hardware or software for this type of test.

Hoffmaster et al.²³ investigated the use of Jamar dynamometry and EMG readings. They found equally low variability in force output (i.e., CV) for both sincere and feigned-weakness conditions but significant differences between sincere and feigned efforts in "amplitude and frequency spectrum of the EMG." The use of the EMG, however, introduces many complexities of its own, including electrode placement and abrasion of the skin to achieve an adequately low resistance.

Arguably the best method of detecting feigned weakness that has been described is that of Stokes et

al.,²⁴ who showed that the difference between peak scores in the five-rung test of grip strength and the peak scores on the REG test was predictive of "sincere" vs. "low" effort. In normal subjects who were instructed to give sincere or low effort and in patients classified as compliant or not compliant by a set of behavioral criteria, compliant subjects showed virtually no difference in the two peak forces, whereas non-compliant subjects showed a substantial difference. A regression model was able to classify normal subjects with 100% accuracy, and patients with 96.6% accuracy (although they did not report whether the errors were in sensitivity or specificity). This approach requires the use of a dynamometer interfaced to a computer. The system also does not involve tests of pinch strength.

Summary

A review of the literature on grip testing by Lechner et al.¹³ concluded that "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 measures for this purpose." Ashford et al.² wrote, "the current protocol for Jamar testing can allow a patient to make a consistently submaximal effort, resulting in a false apparent loss of grip strength." Tredgett et al.⁷ stated, "No single grip strength test has yet been described which can unequivocally detect malingerers."

The approach described below, based on a comparison of standard, unilateral grips and pinches with those during simultaneous bilateral testing, employs "distraction-based testing," described by Waddell et al.²⁵ as "non-emotional, non-surprising and non-hurtful." The high degree of sensitivity to stimuli in the hands and fingers is provided by high-density populations of mechanoreceptors. Fine motor control of force output is facilitated by the small motor units found in the hands and fingers. These two anatomic features provide a biofeedback loop that may be capable of providing very accurate information to test subjects, thereby facilitating submaximal efforts that subsequently go undetected with traditional testing methodology. The methodology used in this investigation was devised to determine whether a new testing protocol could overcome the sensitive biofeedback loop in the upper extremities, a challenge that we believe is responsible for the limited success thus far in accurately classifying effort during strength testing of the hands.

The study reported below was conducted to determine whether a new method of assessment and analysis, static simultaneous bilateral force assessment of hand and pinch strength, is effective in detecting feigned upper extremity weakness while also correctly identifying subjects who are giving a maximal effort.

MATERIALS AND METHODS

Subjects

The 100 subjects in the study were undergraduate students at Millikin University, and volunteers were tested at locations independent of the university. All were, by self-report, asymptomatic for upper-extremity injury or weakness. The mean age was 24 years (range, 18–75 years; SD, 9.7; and 18 subjects were 30 years old or older). Equal numbers of men and women were tested. The research was approved by the Institutional Review Board of Millikin University. The 61 subjects at Millikin University were paid \$15 on completion of two sessions. Other subjects were unpaid volunteers.

The first author tested 63 of the subjects. Other subjects were tested by several of the other authors (RT, TS, and DF).

Apparatus and Procedure

Measurements for this study were taken on newly calibrated instruments. Grip strength testing was performed on a Jamar dynamometer. Pinch strength testing was performed on the Baseline and B&L pinch gauges. Forces were recorded to the nearest 0.45 kg (1 lb) on the Jamar and to the nearest 0.23 kg (0.5 lb) for the pinch gauges. These instruments have been found by Mathiowetz et al.²⁶ and Härkönen et al.²⁷ to be sufficiently accurate for testing patients in a clinical setting.

Subjects were tested in two sessions 1 to 3 days apart. In one session, the subject was instructed to give a maximum effort at all times with both hands. In the other session, the subject was instructed to give an "approximately 50% effort" with one hand at all times and to attempt to give a maximum effort at all times with the other. Order of testing (compliant and noncompliant) was counterbalanced, as was the hand for which the subject was to feign weakness. All conditions were tested equally often with male and female subjects. Subjects held the Jamar dynamometer and Baseline pinch gauge in standard test positions as described by Mathiowetz et al.²⁶

The order of the static strength test trials for hand and pinch strengths was randomized for each subject and for each session. Randomization of the order is designed to minimize the use of "muscle memory" during testing, as a particular test is unlikely to be repeated immediately. The first set of 66 trials, presented in random order, involved unilateral and bilateral tests under instructions to squeeze for 3 to 4 seconds. The unilateral trials included three trials at each of the five Jamar positions and each of the three pinch positions (two-point, three-point, and lateral) for each hand, for a total of 48 unilateral trials. Simultaneous bilateral trials were conducted on each

hand for positions 2, 3, and 4 of the Jamar dynamometer, paired with each of the three pinch positions for the opposite hand. Table 1 lists the types of trial and the number of repetitions of each.

No verbal, auditory, or visual feedback about physical output was given to the subjects during the test. During simultaneous bilateral trials, the evaluator observed the needle of the Jamar dynamometer and lightly held the pinch gauge to determine whether the subjects were exerting force simultaneously. Trials were repeated if the two grips were not initiated and terminated simultaneously.

After all static hand grip and pinch strength measurements were taken, explosive grip trials were performed for each hand at each of five positions on the Jamar dynamometer. The ten explosive trials were also in random order. During the explosive grips, subjects were asked to grasp the instrument and "squeeze one time as hard and as fast as you can" in the instructed-compliant sessions. During instructed-noncompliant sessions, subjects were asked to "squeeze quickly at approximately 50% of your maximum strength level." Trials were repeated if the evaluator believed the duration of the grasp was greater than 0.5 second. Although this may not completely standardize this test, the same problem is inherent in current clinical use of the test.

A software program written by one of the authors (SDStJ) for this protocol randomized the order of trials and instructed the experimenter which trial to conduct. The experimenter then entered the force values. The intent of randomizing the order of the trials was to limit the use of "muscle memory" that may be possible if the same trial is performed repeatedly.

In clinical practice, it is sometimes the case that subjects in any test give equivocal ("gray-zone") results. We chose to classify results as "gray-zone" if the subjects failed one (and only one) of the criteria described below. Results for five subjects in instructed-compliant testing, and for seven in instructed-noncompliant testing, fell into this gray zone. All 12 subjects were retested. Because some criteria were established after data collection was complete, some of the retests were conducted after the completion of initial testing.

RESULTS

The analysis of the data was focused on the detection of noncompliance in testing. To that end, we sought criteria that would correctly classify the maximum number of noncompliant tests while avoiding false alarms (i.e., classifying a compliant test as noncompliant). Criteria for determining noncompliance were deliberately set to minimize false alarms in our sample. That does not guarantee against false alarms in the use of this protocol in clinical testing.

TABLE 1. Trials in Experimental Protocol

Unilateral trials (each hand)	
Jamar grip test:	
Position 1 (3 trials)	
Position 2 (3 trials)	
Position 3 (3 trials)	
Position 4 (3 trials)	
Position 5 (3 trials)	
Pinch test:	
Two-point pinch (3 trials)	
Three-point pinch (3 trials)	
Lateral pinch (3 trials)	
Simultaneous bilateral trials (both hands)*	
Jamar grip test, position 2 (3 trials) with, on other hand:	
Two-point pinch test (1 trial)	
Three-point pinch test (1 trial)	
Lateral pinch test (1 trial)	
Jamar position 3 (3 trials) with, on other hand:	
Two-point pinch test (1 trial)	
Three-point pinch test (1 trial)	
Lateral pinch test (1 trial)	
Jamar position 4 (3 trials) with, on other hand:	
Two-point pinch test (1 trial)	
Three-point pinch test (1 trial)	
Lateral pinch test (1 trial)	

*One set each: Jamar grip test left hand and simultaneous pinch test right hand; Jamar grip test right hand and simultaneous pinch test left hand. Three trials of each pinch test (two-point, three-point, and lateral pinch) are performed in the complete series of tests.

The two-point pinch was not included in developing the criteria, as it produced fairly high rates of "bad" CVs (i.e., of 15% or more) during compliant testing. For instructed-compliant testing for two-finger pinch, 11.5% of the CVs were bad. In contrast, only 4.5% of the three-point and 1.5% of the lateral pinches were bad.

Seven validity criteria were identified, and two approaches to testing validity (the "bell-shaped curve" and a variant of the REG tests) were shown to be ineffective. Each of these is described below. Table 2 reports the seven criteria, along with the mean and standard deviation for instructed-compliant subjects, the z-score of the criterion, and the number of subjects failing each criterion during instructed-compliant and instructed-noncompliant testing.

The z-score is the number of standard deviations that a criterion lies from the mean for that test. For example, the criterion based on the mean number of "bad" change scores (those over 15%) is set at 14% (the criterion is failed if the mean of the changes is 14% or greater). The mean change during instructed-compliant testing was 7.91%, with a standard deviation of 2.44. Thus, the criterion of 14% was 2.5 standard deviations above that mean. In a normal distribution, only 0.6% of cases lie that far from the mean. That somewhat underestimates the true figure, however, as the data are positively skewed. Table 3

TABLE 2. Mean and Standard Deviations for Compliant testing, z Score for the Criterion, and Number of Persons Failing Each Test, on Initial Testing

Criterion	Mean	SD	z for Criterion	Number Failed	
				Compliant	Noncompliant
Number of CVs \geq 15%	0.85	0.97	4.27	0	70
Mean of all CVs	5.91	1.29	2.97	0	77
Number of changes \geq 14%	1.35	1.25	2.92	0	63
Mean of changes	7.91	2.44	2.50	1	77
Mean of certain bilateral CVs	5.38	1.50	3.08	0	73
Number of bilateral CVs \geq 20%	0.07	0.26	7.57	0	67
Bilateral lateral pinch CV \geq 13%	4.60	3.00	2.81	4	62

TABLE 3. Number of Tests That Failed among 100 Subjects in the Initial Noncompliant Testing Sessions

No. of Tests Failed	No. of Subjects Failing
0	0
1	7
2	7
3	11
4	13
5	19
6	13
7	30

reports the number of persons failing on each specific number of criteria.

The database of expected variations and the criteria derived from them are copyrighted by Schapmire and St. James.²⁸

As a check on the possibility that muscular fatigue may occur during testing, six subjects not involved in the main experiment were tested on the performance of 66 maximum voluntary efforts on the Jamar dynamometer in position 2, at a rate of one grip every 20 seconds. The average grip force over the last five trials was only 1 kg (2.2 lb) less than the average of the first five trials. Thus, with the protocol described here, in which the trials alternate between hands as well as between pinching and gripping, muscular fatigue is not a limiting factor.

Criteria Adopted

Criterion 1: Number of "Bad" Coefficients of Variation

This criterion counts the number of CVs that were greater than or equal to 15%. To illustrate, three grips of 34, 40, and 46 kg would have a CV of 15%. Grips of 35, 35, and 45 kg would also have a CV of 15%. Coefficients of variation were obtained in Jamar posi-

tions 1 through 5 and in three-point and lateral pinches, for unilateral testing. For simultaneous bilateral testing, testing was not conducted in Jamar positions 1 and 5. The criterion adopted was that this test is failed if 5 or more CVs, out of 28 possible, were equal to or greater than 15%. With no false positives during instructed-compliant testing, this criterion detected 70 of 100 instructed-noncompliant efforts.

Criterion 2: Average Coefficient of Variation

Coefficients of variation of 15% or more are generally considered problematic, but what of the subject who has consistently elevated CVs but no individual CVs above 15%? A high average across all 28 CVs could also indicate noncompliance. The criterion adopted was that this test is failed if the mean CV is equal to or greater than 9.75%. With no false positives during instructed-compliant testing, 77 of 100 instructed-noncompliant tests failed by this criterion.

Criterion 3: Number of "Bad" Change Scores

This criterion examines the changes between unilateral and bilateral testing—specifically, the absolute percentage of change— $(| \text{bilateral-unilateral} | / \text{unilateral}) \times 100$. It is based on the number of changes above 15% out of the ten tests (tests at Jamar positions 2 through 4 and lateral and three-point pinch tests for each hand). The criterion adopted was that this test is failed if 5 or more changes, out of 12 possible comparisons, are equal to or greater than 15%. With no false positives during instructed-compliant testing, 63 of 100 instructed-noncompliant tests failed by this criterion.

Criterion 4: Average Percentage Change

Again, the possibility exists that a person might have few "bad" change scores by the 15% criterion and yet have an average change score that was suspect. The criterion adopted was that this test is failed if the mean of the changes is equal to or greater than

14%. By the criterion chosen, there was a single false-positive result during instructed-compliant testing (which was correctly classified as compliant on retesting). In instructed-noncompliant testing, 77 of 100 tests failed by this criterion.

Criterion 5: Count of Simultaneous Bilateral Coefficients of Variation over 20%

This criterion was based on the observation that CVs are consistently higher during simultaneous bilateral testing for instructed-noncompliant subjects. In instructed-compliant testing, only seven subjects had even a single CV in excess of 20%, whereas in the instructed-noncompliant testing, 67 subjects had two or more CVs in excess of 20%. The criterion adopted was that this test is failed if 2 or more of the 12 simultaneous bilateral CVs are equal to or greater than 20%. Thus, with no false positives during instructed-compliant testing, 67 of 100 instructed-noncompliant tests failed by this criterion.

Criterion 6: Mean Coefficients of Variation for Selected Simultaneous Bilateral Tests

This criterion was based on the examination of the differences between CVs obtained on the same subjects in instructed-compliant and instructed-noncompliant conditions. The largest differences were in the results of the simultaneous bilateral tests, and of those, the results for lateral pinch and Jamar position 3 and 4 grip tests showed the consistently largest differences across both hands. Thus, these tests are likely to give large CVs in noncompliant subjects and small CVs in compliant subjects. The mean of these six CVs forms the basis of this criterion. The criterion adopted was that this test is failed if the mean CV of those six tests is equal to or greater than 10%. With no false positives during instructed-compliant testing, 73 of 100 instructed-noncompliant tests failed this criterion.

Criterion 7: Coefficients of Variation for Lateral Pinches During Simultaneous Bilateral Tests

In our data, the difference between lateral pinch CVs for instructed-compliant and instructed-noncompliant testing was particularly large. This led to the suggestion that large CVs in this position could be a sensitive indicator of noncompliance. The criterion adopted was that this test is failed if either of the two lateral pinch CVs during simultaneous bilateral testing is equal to or greater than 13%. Because the distribution of values on this criterion during instructed-compliant testing is positively skewed, the criterion adopted (which is 2.81 SDs above the mean for compliant testing) did lead to four false-positive results during instructed-compliant testing (all of which were correctly classified as compliant on

retesting). However, 62 of the 100 instructed-noncompliant subjects failed this criterion. Two of those passed on all other criteria in the instructed-noncompliant testing and would have been incorrectly classified as compliant had they not failed this criterion. Both failed two or more criteria on retesting and thus were correctly categorized.

Criteria Rejected

Explosive Grip Strength

The rationale for explosive tests (such as the REG test proposed by Lister in Hildreth et al.¹⁵) is that the explosive grip strength should be less than the maximum phasic grip strength for the same position on the Jamar dynamometer and that it is more difficult to produce a submaximal contraction during explosive testing. As a result, if a person produces a grip strength in the explosive test that is greater than the maximum of their three phasic trials, according to this rationale they are almost certainly noncompliant. The mean number of "positive" explosive results (explosive force exceeded maximum regular force) was 1.98 (of a possible 10) for instructed-compliant subjects and 3.41 for instructed-noncompliant subjects. During instructed-compliant testing, 71% of subjects had at least one "positive" explosive grip.

Adopting the most accurate criterion of 6 "positive" explosive grips (of 10) as indicating noncompliance, this criterion correctly classified only 21 of the instructed-noncompliant tests while incorrectly classifying 10 of the instructed-compliant tests as noncompliant.

As noted above, the REG technique can lead to false readings on the Jamar dynamometer because of sudden contact between the hand and the Jamar. To avoid this problem, we used isolated, single explosive grips rather than the REG technique. Nevertheless, this criterion had low sensitivity and an unacceptably high rate of false positive classifications.

Changes in the "Bell-shaped Curve"

We attempted to numerically define deviation from the expected curve of force as a function of position on the Jamar dynamometer. Force is almost always highest at position 2 or 3. Therefore, we defined that deviation as occurring when the weakest grip was in position 2 or 3 during unilateral testing. During simultaneous bilateral testing, that deviation was defined as occurring when position 3 was weaker than both positions 2 and 4 by at least 10%. By this numeric definition, only 18 subjects failed this criterion during instructed-noncompliant testing. Furthermore, 13 of 18 subjects who failed this criterion during instructed-noncompliant testing also failed at least four other criteria. One subject failed

the criterion during instructed-compliant testing. Thus, this criterion had low sensitivity and added almost nothing to the ability to detect feigned weakness.

Summary

In initial testing, 95 of the 100 instructed-compliant subjects passed all seven criteria, while the remaining 5 each failed a single criterion. For the instructed-noncompliant subjects, 93 failed two or more criteria, whereas 7 failed a single criterion. The validity of the test results in terms of compliance was considered equivocal ("gray-zone") for subjects failing a single criterion. Failure of a single criterion was classified as "gray zone" because this was the only condition in which there was any overlap of the two groups. All subjects with "gray-zone" results were then retested in the appropriate condition (instructed-compliant or instructed-noncompliant). On retesting, instructed-noncompliant subjects were instructed to feign weakness with the same hand used in the original testing.

On retesting, six of the seven instructed-noncompliant subjects failed two or more validity criteria and thus were now correctly classified as noncompliant. One subject in the instructed-noncompliant testing completely eluded the validity criteria. All five subjects in the instructed-compliant condition passed all criteria on retesting and so were correctly classified as compliant.

The mean number of failed validity criteria in the instructed-compliant sessions was 0.05 (SD, 0.22), while the mean in the instructed-noncompliant sessions was 4.89 (SD, 1.92).

The mean number of validity criteria failed in instructed-noncompliant testing was similar for men (4.70 criteria failed) and women (5.08), for those feigning weakness in the right (4.98) and left (4.71) hands, and for those tested in the noncompliant condition in the first (4.97) or second (4.80) session. None of these differences approached significance with an independent-samples *t* test (all values of *t* less than 1.0). During noncompliant testing, subjects produced a force of 73% of compliant force, based on an average of all grips and pinches.

A chi-square test of equality of proportions was used to compare—by sex, hand of feigned weakness, and whether the noncompliant testing was in the first or second session—the proportion of subjects failing each validity criterion. None of the tests approached significance by conventional standards.

Other Considerations

The average duration of each trial was 17 seconds (approximately 3-second grip and 14-second rest), as recorded during testing of the first 24 subjects in this experiment. The final recommended protocol, which

omits the explosive grip testing, consists of 66 tests and can be completed in approximately 20 minutes by a practiced clinician.

Muscular fatigue was not likely to be a factor that would affect the outcome of the study. The randomized order of testing resulted in activity alternating between hands and between activities (gripping and pinching). Randomization and the 17 seconds between trials should have allowed for adequate replenishment of substrates and removal of waste products of localized cellular metabolism.

DISCUSSION

In this study, using empirically derived criteria, 94% of subjects were unambiguously correctly classified as compliant or noncompliant in grip strength testing. Of the remaining 6% of "gray-zone" results, all but one were correctly classified on retesting. In this study, sensitivity was 99% and specificity was 100%. Overall accuracy was 99.5%. Using a phi coefficient, the reliability of the test, $\phi_{(198)}$, was calculated to be 0.99 ($p=0.000$).

In the clinical setting, equivocal test results usually justify re-assessment. That would be our recommendation for assessing hand and pinch strengths using this methodology, provided that only a single validity criterion is failed. Failure of two or more criteria occurred exclusively during instructed-noncompliant testing.

The randomization of the order in which the tests were performed added to the distraction of simultaneous bilateral activity and is likely to have been an important factor in eliminating a sensory engram that would allow for the consistent reproduction of submaximal force.

In part, the controversy over the use of the CV centers on the use of this statistic as a stand-alone evaluation tool. This concern has merit, particularly if judgments about effort are made solely on the basis of the results obtained in one of five positions on the testing apparatus. Quite clearly, it would be inherently unfair, and statistically indefensible, to classify effort on the results of a limited number of data sets. In the clinical setting, the classification of effort should be based on a pattern of performance, not a single set of numbers.

In our protocol, consisting of multiple data sets coupled with the distraction of randomized unilateral and simultaneous bilateral activity, the average CV proved to be an effective discriminator between "sincere" and "feigned weakness" sessions. To assess validity of effort, the authors believe that this global use of the CV in a distraction-based test overcomes the deficiencies of the CV as described in previous studies.

The explosive grip testing and a numerically defined test of deviation from the "bell-shaped

curve" lacked sensitivity and specificity and are not recommended for use in clinical testing.

One concern in testing a clinical population is that pain could produce an artificially high CV. A reasonable scenario is that a person might squeeze as hard as possible on the first try, then limit the effort to below the threshold of pain on subsequent tests. This practical problem can be overcome by having the subject make a grip, a three-point pinch, and a lateral pinch with the affected hand before beginning testing. This could easily be done in the course of familiarizing the patient with the equipment. However, we would submit that the pain response would not be likely to account for differences between results of unilateral and bilateral trials in the simultaneous testing.

No separation between hand grip and pinch strengths was made in this analysis. It is reasonable to expect some pathologic conditions to affect only hand grip or only pinch strength. However, weaknesses in pinch and grip strength may be present with some pathologic conditions. The present approach would produce results that could objectively identify the area of weakness in compliant patients. In persons who are consciously attempting to control test results, a global (as opposed to focused) weakness is likely to be present, if the findings of Menard et al.²⁹ apply to upper-extremity patients. Therefore, separate validity criteria for hand and pinch strengths are not necessary or practical in a clinical setting. Although we did not use two-point pinch in developing the validity criteria described above, the measure of this specific force may be useful diagnostically and therefore remains a part of the protocol. A valid effort in the other parts of the test would presumably indicate that the two-point pinch results are valid as well.

An advantage of the protocol described in this study is ease of administration. The equipment is widely used in the clinical setting. While a clinician can perhaps contrive to perform the procedure without supporting software, the most practical solution, the one used in this study, involves use of the software program.²⁸ Properly calibrated, the Jamar dynamometer and the Baseline and B&L pinch gauges provide acceptable levels of accuracy for upper extremity evaluation. Expensive equipment, such as isokinetic devices or computerized equipment that records time-force data, is not required for this protocol.

One strength of the method we recommend is the multiple criteria. Failure on two or more criteria is thus nearly certain evidence of feigned weakness. Failure on one criterion is equivocal and justifies retesting.

One of the strengths of the methodology described in this study is its high degree of accuracy coupled with a low rate of "equivocal" test results. In the clinical setting, findings of true strength deficits as

revealed in this test protocol could justifiably be used to support recommendations for testing or treatment. Findings of feigned weakness would justify a request to re-evaluate the patient's case and make appropriate recommendations for testing, treatment, or non-medical case management.

Further specifications of the statistical criteria used for each test, along with norms based on our sample of subjects tested under conditions of known compliance and non-compliance, are given in Schapmire and St. James.²⁸

LIMITATIONS

An obvious limitation of the present study is the restriction of subjects to an asymptomatic and (mostly) young sample. To validate the protocol with injured patients, a second study is currently being conducted to apply the same protocol to patients with known injury or pathology. One requirement for such a study is that the subjects not be eligible for compensation, so that they can be presumed to comply with the instruction to produce a sincere or insincere effort.

A limitation of the protocol developed by this study is the time required to score the results. As noted in the review of numerous articles, easily scored computations used in other studies have not been able to reliably detect feigned weakness. This underscores the complexity of assessing validity of effort—that complex calculations are required to do so adequately. Software to permit computerized entry of the strength tests and automatic calculation of the criteria is commercially available.²⁸

All equipment was initially tested for calibration, but periodic recalibration was not undertaken throughout the study. The absence of significant false-negative and false-positive results is offered as evidence that the equipment did not fall out of calibration over time. Admittedly, this is an assumption, rather than hard data.

It is impossible to reproduce the grip of a true feigner, as no one knows how feigners actually feign their grip efforts. The method of asking the subjects to give a 50% effort was selected not to be an exact duplication of feigned effort but to ensure a less-than-sincere effort.

REFERENCES

1. Robinson ME, MacMillan M, O'Connor P, Fuller A, Cassisi JE. Reproducibility of maximal versus submaximal efforts in an isometric lumbar extension task. *J Spinal Disord.* 1991;4:444-8.
2. Ashford R, Nagelburg S, Adkins R. Sensitivity of the Jamar dynamometer in detecting submaximal grip effort. *J Hand Surg.* 1996;21A:402-5.
3. Dvir Z. Coefficient of variation in maximal and feigned static and dynamic grip efforts. *Am J Phys Med Rehabil.* 1999;78:216-21.
4. Fairfax AH, Balnave R, Adams R. Variability of grip strength

- during isometric contraction. *Ergonomics*. 1995;8:1819-30.
5. Niebuhr B, Marion R. Voluntary control of submaximal grip strength. *Am J Phys Med Rehabil*. 1990;69:96-101.
6. Simonsen J. Coefficient of variation as a measure of subject effort. *Arch Phys Med Rehabil*. 1995;76:516-20.
7. Tredgett M, Pimble LJ, Davis TR. The detection of feigned hand weakness using the five position grip strength test. *J Hand Surg*. 1999;24B:426-8.
8. Shechtman O. Using the coefficient of variation to detect sincerity of effort of grip strength: a literature review. *J Hand Ther*. 2000;13:25-32.
9. Shechtman O. The coefficient of variation as a measure of sincerity of effort of grip strength, Part I: the statistical principle. *J Hand Ther*. 2001;14:180-7.
10. Shechtman O. The coefficient of variation as a measure of sincerity of effort of grip strength, Part II: sensitivity and specificity. *J Hand Ther*. 2001;14:188-94.
11. Stokes HM. The seriously uninjured hand. *J Occup Med*. 1983;25:683-4.
12. Niebuhr B, Marion R. Detecting sincerity of effort when measuring grip strength. *Am J Phys Med*. 1987;16-24.
13. Lechner D, Bradbury S, Bradley L. Detecting sincerity of effort: a summary of methods and approaches. *Phys Ther*. 1998;78:867-98.
14. Hamilton A, Balnave R, Adams R. Grip strength testing reliability. *J Hand Ther*. 1994;7:163-70.
15. Hildreth D, Breidenbach W, Lister G, Hodges A. Detection of submaximal effort by use of the rapid exchange grip. *J Hand Surg*. 1989;14A:742-5.
16. Joughin K, Gulati P, MacKinnon S, et al. An evaluation of rapid exchange and simultaneous grip tests. *J Hand Surg*. 1993;18A:245-52.
17. Taylor C, Shechtman O. The use of the rapid exchange grip test in detecting sincerity of effort, part I: administration of the test. *J Hand Ther*. 2000;13:195-202.
18. Shechtman O, Taylor C. The use of the rapid exchange grip test in detecting sincerity of effort, Part I: validity of the test. *J Hand Ther*. 2000;13:203-10.
19. Chengalur S, Smith GA, Nelson RC, Sadoff AM. Assessing sincerity of effort in maximal grip strength tests. *Am J Phys Med Rehabil*. 1990;69:148-53.
20. 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.
21. Gilbert JC, Knowlton RG. Simple method to determine sincerity of effort during a maximal isometric test of grip strength. *Am J Phys Med*. 1983;62:135-44.
22. Mitterhauser MD, Muse VL, Dellon AL, Jetzer TC. Detection of submaximal effort with computer-assisted grip strength measurements. *J Occup Environ Med*. 1997;39:1220-7.
23. Hoffmaster E, Lech R, Niebuhr B. Consistency of sincere and feigned grip exertions with repeated testing. *J Occup Med*. 1993;35:788-94.
24. Stokes HM, Landrieu KW, Domangue B, Kunen S. Identification of low-effort patients through dynamometry. *J Hand Surg*. 1995;20A:1047-56.
25. Waddell G, McCulloch JA, Kummel E, Venner RM. Nonorganic physical signs in low-back pain. *Spine*. 1980;5:117-25.
26. Mathiowetz V, Weber K, Volland G, Kashman N. Reliability and validity of grip and pinch strength evaluations. *J Hand Surg*. 1984;9A:222-6.
27. Härkönen R, Harju R, Alaranta H. Accuracy of the Jamar dynamometer. *J Hand Ther*. 1993;6:259-62.
28. Schapmire D, St. James JD. X-RTS Hand Grip Assessment: Instructions and Norms. Hopedale, Ill.: X-RTS Software Products, 2001.
29. Menard MR, Cooke C, Locke SR, Beach GN, Butler TB. Pattern of performance in workers with low back pain during comprehensive motor performance evaluation. *Spine*. 1994;19:1359-66.