Validity and reliability of sincerity test for dynamic trunk motions

S. A. FERGUSON*, S. GALLAGHER and W. S. MARRAS

Biodynamics Laboratory, The Ohio State University, Columbus, OH, USA

Abstract

Purpose: Marras *et al*. developed a technique to evaluate sincerity of effort during dynamic trunk motion performance. The validity and reliability of the technique have not been evaluated. Therefore, the objective of this study was to first determine whether or not a sincerity of effort measure correctly identified those giving a sincere effort in a blinded randomized control trial and second to quantify inter–rater and test–retest reliability.

Methods: This article reports the findings of a two phase study. In phase one, the blinded evaluation, participants were randomly assigned to either a sincere or insincere performance condition. An examiner tested participants without knowledge of the participant's group membership. In the second phase, two examiners evaluated each participant twice to quantify inter–rater and test–retest reliability.

Results: In the blinded phase the specificity was 100% and sensitivity was 90% for identifying sincere and insincere effort, respectively. Phase two results showed no significant difference in probability of sincere effort between raters or between testing sessions.

Conclusion: A performance criterion that accurately identifies sincere vs insincere group membership during functional evaluations was identified. There were no significant differences between raters or between testing sessions. These findings indicate that this test is reliable and possesses good predictive validity in assessing sincerity of effort.

Introduction

Despite the vast amount of research on prevention of low back disorders (LBD), these disorders continue to impose an enormous burden on society.[1,2] Epidemiological studies have shown that the dynamic characteristics of the job are good indicators of the risk of low back injury.[3] A low back functional assessment tool that quantifies the dynamic low back capacity of an injured worker may provide a mechanism to directly compare functional capability to dynamic job demands. A quantitative functional performance tool could measure the severity of the initial injury, measure any improvement from treatments, and potentially evaluate the readiness to return to work.[4–7] Furthermore, dynamic functional assessment results may be compared to dynamic characteristics of job demands in order to select light duty job.

Quantitative performance assessment measures (e.g. strength or kinematic measures) can provide detailed information on the patient's musculoskeletal status. However in order for these assessments to be useful and reliable, the patient must perform a true or sincere effort. Marras *et al*.[5,8] developed a dynamic functional assessment tool as well as technique to assess whether or not a person is performing a true effort (sincere) or an effort that exacerbates the illness behaviour (insincere) during a quantitative functional performance evaluation. Marras *et al*.[5,8] found that the phase plane motions from a sincere performance were more repeatable or consistent than phase plane motions from an insincere effort. The hypothesis from these authors' states that during a sincere motion the musculoskeletal system relies on a well developed motor programme to recruit the trunk musculature necessary to create the motion. During an insincere effort the motor programme is mediated and becomes less consistent. A model characterizing the probability of sincere effort was developed to ensure that functional performance evaluation data was of high quality and could be trusted to evaluate LBDs.

In order for such sincerity tests to be used in clinical practice they must be validated through randomized clinical trials. These tests can serve two purposes. First, the validity of such tests can be evaluated by its ability to accurately identify group membership as sincere and insincere performance during a functional evaluation.[9] In this case, sensitivity would be the ability to correctly identify those giving an insincere effort and specificity would be correctly identifying those giving a

* Author for correspondence; e-mail: ferguson.4@osu.edu

sincere effort. Thus, there is a need to validate the sensitivity and specificity of the tool from a single evaluation.

Second, the reliability of a measure must be verified.[10] Two specific types of reliability relevant to the development of a medical assessment tool are inter–rater reliability, which is consistency between observers and test–retest reliability, which addresses stability of ratings over time.[10, 11] The objective of this study was to evaluate the validity and reliability of the technique described by Marras *et al*.[8] The specific purposes of the study are: (1) to determine whether the model effectively discriminates between those providing a sincere and insincere effort; and (2) to assess the reliability the probability of sincere effort protocol.

Methods

APPROACH

In order to assess the validity and reliability of the probability of sincere effort a two-phase study was developed. Phase one consisted of a blinded randomized clinical trial to evaluate whether or not the probability of sincere effort accurately identified the participants providing a sincere effort during a functional performance evaluation. The dynamic functional performance evaluation developed by Marras *et al*.[8] was performed using the lumbar motion monitor. The probability of sincere effort was calculated from the model using motion characteristics from the functional performance evaluation. This probability can be used to differentiate between participants giving a sincere vs insincere effort and can be dichotomized at different cut-off points to evaluate specificity and sensitivity. Phase two of the study was developed to evaluate inter–rater reliability as well as test–retest reliability. The severity of symptoms in low back patients vary over time, which may influence functional performance consistency therefore, it was decided that reliability should be evaluated with asymptomatic participants. This two-phase study was designed to assess reliability as well as validate the 'probability of sincere effort' model. Further validation using clinical patients will be necessary however the current study must show positive results to justify such an investment.

PARTICIPANTS

Twenty asymptomatic participants with no previous history of low back pain were recruited for the study. There were 10 male and 10 female participants. Males had an average (standard deviation) standing height of 178.5 cm (6.1), weight of 88.9 kg (27.3) and age of 22.0 (2.1). Females had an average standing height of 165.0 cm (6.1), weight of 64.5 kg (14.6) and age of 21.7 (0.95).

EXPERIMENTAL DESIGN

Phase one was a blinded randomized clinical trial. Researcher 1 randomly assigned the participants to the sincere group or the insincere group. Researcher 2 who performed the functional performance evaluations was blinded to the group assignment given each participant.

The second phase was a reliability study to evaluate differences between raters and between testing sessions. The independent variables were rater and trial. The dependent measure was the probability of sincere effort. In this study, two raters evaluated each participant twice, for a total of four evaluations per participant. The trials were one week apart. Each trial took place on the same day of the week at the same time of day.

In both phases the dependent measure was probability of sincere effort.[8] In order to calculate probability of sincere effort four functional performance tasks were required. The tasks include: (1) controlled sagittal flexion and extension; (2) uncontrolled sagittal flexion and extension; (3) uncontrolled axial twisting about the waist; and (4) uncontrolled lateral bending. The controlled sagittal task required the participant to flex and extend their trunk while maintaining their twisting motion within $+/-2^\circ$ of zero in the transverse plane. If the participant failed to maintain control within the tolerance the trial was repeated. In the three uncontrolled tasks participants were instructed to generate motion in a given plane with no off-plane tolerances. These four tasks were completely randomized at each testing session.

PROCEDURE

Participants were informed of the length of the study and number of testing sessions according to an approved Institutional Review Board protocol. Participants were informed that they could drop out at any time during the testing sessions. Typically testing sessions were not more than 20 minutes. The data collection time of each exertion was 8 seconds.

Upon arrival at each session the participant signed an IRB consent form. The lumbar motion monitor (LMM) was placed on the participant.[12] Each participant was instructed to 'cross your arms in front of you and stand with your feet shoulder width apart'. The participants performed four tasks including flexing and extending (to upright standing) the trunk, twisting clockwise and counterclockwise, bending side to side, and flexing and extending while controlling the twisting position within $+/-2^{\circ}$ of zero displayed on the computer screen. The instructions for the tasks were to move 'as fast as you can comfortably'.

The blinded randomized clinical trial examined the same four trunk motions; however, subjects were given different instructions depending on their random group assignment, which was established upon their arrival. The participants in the sincere group were instructed to perform all the tasks 'as fast as you can comfortably'. The insincere group was instructed to 'perform the tasks as if you had severe low back pain'. To ensure that the participant understood what region of the back was included in low back pain these participants were shown the L1 – L5 region of their back. In all cases, the participant was told not to discuss the instructions with the examiner. Thus, the examiner was blinded to the instructions provided to the participant regarding the sincerity of effort.

APPARATUS

The LMM was used to measure trunk motion in all three planes of the body. The device is an exoskeleton attached to harness worn around the pelvis and thorax. The LMM measures instantaneous changes in position between the pelvis and thoracolumbar region. A detailed account of the calibration may be found in Marras *et al.*[12] The LMM signal was collected at 60 Hz and stored on a microcomputer and analysed later to derive velocity, acceleration and jerk.

In the controlled sagittal bending condition the microcomputer provided a display illustrating the tolerance. The tolerance was +/− two degrees in the transverse plane. If the participant went outside the control zone during the test then the trial was automatically terminated and repeated.

DATA ANALYSIS

Custom software was used to derive the position as a function of time in the sagittal, frontal and transverse planes of the body.[8] The position data was conditioned to determine velocity, acceleration and jerk. The probability of sincere effort was calculated using a discriminant function model that was developed from a previous study.[8] In the previous study participants were asked to perform both a sincere and insincere effort of the same tasks. The data from the previous study was used for model development.[8] The data from the current study was treated as test data in the model. The motion parameters used to calculate probability of sincerity were: (1) standard deviation of position from the lateral bending task; (2) coefficient of variation from acceleration from the lateral bending task; (3) standard deviation of velocity from the sagittal uncontrolled task; (4) coefficient of variation for the velocity-acceleration phase plane from the twisting task; and (5) coefficient of variation for the acceleration-jerk phase plane from the twisting task.

Analysis of the blinded randomized clinical trial included descriptive statistics and an evaluation of the distribution of the data. A *t*-test was performed to evaluate whether there was a significant difference in model estimates of the probability of sincere effort between those performing sincerely vs insincerely. Finally, an evaluation of the cut-off point was performed to determine the best classification for a categorical measure sincere and insincere effort in terms of specificity and sensitivity.

An analysis of variance (ANOVA) was performed for phase two. The analysis was performed with SAS (Cary, NC) software on an IBM mainframe computer. The ANOVA was performed using rater, rater nested within trial and subjects as independent measures.

Results

All participants completed both the reliability study and blinded study.

PHASE ONE

The mean probability of sincerity for the sincere group was 0.87 with a standard deviation of 0.09. The insincere group on the other hand had a mean probability of 0.40 with a standard deviation of 0.21. A *t*-test was performed and the unequal variance *p*-value was used to evaluate whether or not a significant difference occurred between the two groups. The results showed that the sincere group had significantly higher probability of sincerity than the insincere group with a *p*-value of 0.001. There was no significant difference between the sincere and insincere groups for age, height and weight.

Figure 1 shows a scatter plot of the probability of sincerity for each subject by group. It clearly illustrates that the sincere group consistently has a probability of sincerity around 0.9, whereas the insincere does not have a consistent probability of sincerity. This illustrates that the variability in the insincere group is greater than the sincere group. In order to interpret the data figure 2 was

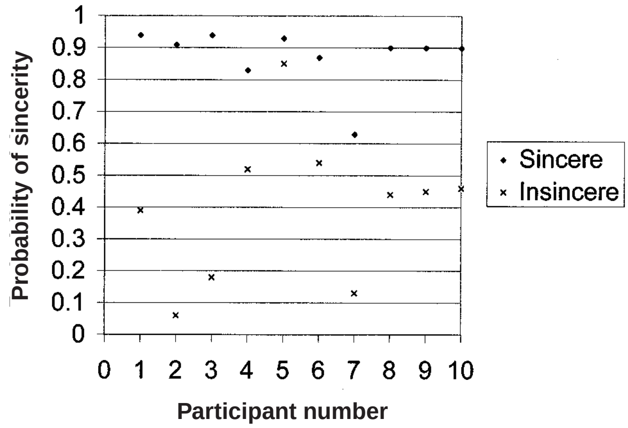

Figure 1 Probability of sincerity for each participant by group.

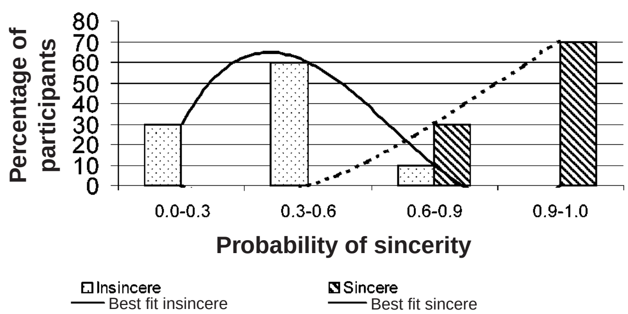

Figure 2 Percentage of participants by probability scores and effort. Subject responses divided into probability ranges (0-0.3, 0.3-0.6, 0.6-0.9, 0.9-1.0).

created, which shows the percentage of the population as a function of probability of sincerity group as well as a best fit polynomial distribution to the data for the sincere and insincere groups. The figure shows that 70% of the sincere participants had a probability of 0.9 or above. Thirty percent of the insincere population had a sincerity probability of less than 0.30. Sixty per cent of the insincere group had a sincerity probability between 0.30 and 0.6 and only 10% had a score over 0.6. The best fit polynomials illustrates the small overlap between the two distributions indicating the good distinction between sincere and insincere effort using this evaluation.

Figure 2 can be used as a guide for selecting the cut-off point for dichotomizing the probability of sincere effort. The cut-off point of 0.6 would result in a specificity of 100% and a sensitivity of 90%. The high specificity and sensitivity clearly indicates that the probability of sincere effort accurately quantifies the quality of effort being put forth by the participant. These results indicate that the measure of 'probability of sincere effort' can validly discriminate between participants based on the nature of their efforts.

Table 1 Mean probability of sincere effort as a function of rater and trial

Rater	*Trial*	*Mean probability of sincerity (STD)*
1	1	0.812 (0.11)
1	2	0.845 (0.09)
2	1	0.810 (0.10)
2	2	0.790 (0.15)

PHASE TWO

Table 1 lists the rater, trial and mean probability of sincerity for the population and standard deviation for probability of sincerity. The ANOVA showed no significant difference between raters. The *p*-value for the rater variance was 0.5839. Trial was nested within rater and showed no significant effect with a *p*-value of 0.2693. As expected there were significant differences among the subjects with a *p*-value of 0.0047. Phase two results indicate that the probability of sincere effort protocol is reliable both between raters and across trials.

Discussion

Understanding low back disorders and low back injury recovery is a complex issue. There are numerous factors that may influence recovery including psychological, psychosocial, physical workplace demands, personal factors and the often overlooked issue of the definition of recovery.[13] A quantitative functional performance measure may provide one piece of the complex puzzle for understanding the musculoskeletal status of individuals with LBDs. However, a functional performance evaluation is only useful if the performance measure is reliable. The probability of sincere effort was developed to ensure high quality data during a quantitative functional performance evaluation.

Functional performance evaluations may be used in several ways. First, it may quantify the severity of the injury. Second, it may quantify the amount of improvement from treatment. Finally, functional performance may be matched with job demands to determine whether or not a person should return to work or return to a light duty job. Furthermore, it may serve as a guide for selecting the criteria for light duty. Thus if a person is providing an insincere effort during a functional performance evaluation practitioners may be erroneously assigning the worker to a light duty job.

Main and Waddell[14] discussed the misuse and misinterpretation of 'non-organic signs' therefore it is cautioned that this procedure is not intended to identify malingerers but rather indicates the quality of functional performance data. The probability of sincerity was developed to be used in conjunction with the 'probability of normal'[5] which is used to evaluate the musculoskeletal status of a low back pain patient. The concept is that if both probabilities are low then the patient should be re-tested. If on the other hand probability of sincerity is high and the probability of normal is low this would indicate the patient gave a sincere effort and truly has low back functional impairment.

The probability of sincere effort is a continuous variable and may provide the most useful information to the practitioner in that form. The validity of a screening tool is its ability to correctly identify true cases (sensitivity) and false cases (specificity).[9] One of the goals of the study was to validate the probability of sincere effort, therefore it is necessary to examine sensitivity and specificity. In order to evaluate specificity and sensitivity, a cut-point must by chosen to dichotomize the continuous measure. A cut-off point of 0.6 provides a specificity of 100% and a sensitivity of 90%. This cut-off point provides excellent sensitivity and specificity, therefore validating that indeed the probability of sincere effort can identify those performing a true effort. In practice, it may be better to use the probability of sincere effort as a continuous variable. However, to ensure high quality data for interpretation on functional performance evaluations, it would be recommended that any performance with a probability of sincere effort less than 0.6 should be repeated.

Jay *et al.*[15] found greater variability in those performing an insincere exertion. It is hypothesized that those performing a sincere effort are calling a central set[15] or motor programme to perform the task where as those performing an insincere task are over-riding the central set with a new recruitment programme. Furthermore, the degree to which the participant over-rides the central set may influence the probability of sincerity. This may be one explanation for the greater variability found in the insincere group compared to the sincere group.

The inter–rater and test–retest reliability findings indicate no significant difference between examiners or visits. This shows that the stability of the probability of sincere effort remains uniform between observers and trials. Thus, the findings of this study complete one part of a complex puzzle. Now it is known how to quantitatively determine sincere and insincere trunk motion performance in asymptomatic controls.

There are several limitations in this study. First, the sample population was asymptomatic. Asymptomatic participants were chosen over low back pain patients because low back pain patients may have high variability in functional performance from day to day due to their symptoms. Since one of the goals of the study was to evaluate the reliability of the test over time, it was decided that a low back population would not be appropriate. The next phase of research will include a blinded randomized control study with low back pain patients. Second, the population was young, predominantly in their twenties. The probability of sincerity is normalized by age and gender, therefore it was hypothesized that age would not affect the results of the current study.

Conclusions

The study showed that a model developed to assess sincerity of effort in performing trunk motion was able to discriminate with high sensitivity and specificity between those giving a sincere effort and those performing insincerely. There were no significant differences between raters or trials in determining the probability of sincere effort by the participants.

References

1 Bigos S, Battie M. The impact of spinal disorders in industry. In: JW Frymoyer (ed.) *The Adult Spine: Principles and Practice*, second edition. Philadephia: Lippincott-Raven, 1997; 151–161.

2 Frymoyer J, Durett C. The economics of spinal disorders. In: J Frymoyer (ed.) *The Adult Spine: Principles and Practice*, second edition. Philadephia: Lippincott-Raven, 1997; 143–150.

3 Ferguson S, Marras W. A review of low back disorder surveillance measures and risk of factors. *Clinical Biomechanics* 1997; **12**: 211–226.

4 Frymoyer J. Quality: an international challenge to the diagnosis and treatment of disorders of the lumbar spine. *Spine* 1993; **18**: 2147–2152.

5 Marras W, Ferguson S, Gupta P *et al*. The quantification of low back disorder using motion measures: methodlogy and validation. *Spine* 1999; **24**: 2091–2100.

6 Teasell R, Harth M. Functional restoration. Returning patients with chronic low back pain to work: revolution or fad? *Spine* 1996; **21**: 844–847.

7 Waddell G, Allan D, Newton M. Clinical evaluation of disability in low back pain. In: J Frymoyer (ed.) *The Adult Spine: Principles and Practice*, second edition. Philadelphia: Lippincott-Raven, 1997; 171–183.

8 Marras W, Lewis K, Ferguson S, Parnianpour M. Impairment magnification during dynamic trunk motions. *Spine* 2000; **25**: 587–595.

9 Andersson GBJ. Sensitivity, specificity, and predictive value a general issue in the screening for disease and in the interpretation of diagnostic studies in spinal disorders. In: JW Frymoyer (ed.) *The Adult Spine: Principles and Practice*, first edition. New York: Raven Press, 1991; 277 – 287.
10 Johnston M, Keith R, Hinderer S. Measurement standards for interdisciplinary medical rehabilitation. *Archives of Physical Medicine and Rehabilitation* 1992; **73**: S3 – S23.
11 Allen M, Yen W. *Introduction to Measurement Theoy*. Monterey: Brooks/Cole Publishing Co., 1979; 72 – 117.
12 Marras W, Fathallah F, Miler R, Davis S, Mirka G. Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *International Journal of Industrial Ergonomics* 1992; **9**: 75 – 87.
13 Ferguson S. Quantification of low back pain recovery using biomechanical, symptom, activities of daily living and work status measures. The Ohio State University, Columbus 1998.
14 Main C, Waddell G. Behavioural response to examination a reapraisal of the interpretation of 'nonorganic signs'. *Spine* 1998; **23**: 2367 – 2371.
15 Jay M, Lamb J, Watson R, Young I, Feeron F, Alday J, Tindall A. Sensitivity and specificity of the indicators of sincere effort of the EPIC lift capacity test on a previously injured population. *Spine* 2000; **25**: 1405 – 1412.
16 Horak F, Diener H. Cerebellar control of postural scaling and central set in stance. *Journal of Neurophysiology* 1994; **72**: 479 – 493.