

# The importance of diagnostic test parameters in the interpretation of clinical test findings: The Prone Hip Extension Test as an example

Paul Bruno, BHK, DC, PhD\*

---

*The use of diagnostic tests is a crucial aspect of clinical practice since they assist clinicians in establishing whether a patient has or does not have a particular condition. In order for any clinical test to be used most appropriately, it is essential that several parameters be established regarding the test and that these are made known to clinicians to inform their clinical decision making. These include the test's sensitivity, specificity, predictive values, and likelihood ratios. This article reviews their importance as well as provides an illustrative example that highlights how knowledge of the parameters for a given test allows clinicians to better interpret their test findings in practice.*  
(JCCA 2011; 55(2):69-75)

---

*L'utilisation de tests diagnostiques est un aspect crucial de la pratique clinique puisqu'ils aident les cliniciens à déterminer si le patient présente ou ne présente pas un état pathologique particulier. Pour que tout essai clinique soit utilisé de façon appropriée, il est essentiel que plusieurs paramètres soient établis en rapport au test, et que ceux-ci soient portés à la connaissance des cliniciens afin d'appuyer leurs prises de décisions cliniques. Cela comprend la sensibilité du test, la particularité, les valeurs prédictives et les rapports de vraisemblance. Cet article examine leur importance et fournit un exemple indicatif qui illustre comment la connaissance des paramètres pour un test donné permet aux cliniciens de mieux interpréter leurs résultats de test dans la pratique.*  
(JCCA 2011; 55(2):69-75)

KEY WORDS: diagnostic test, clinical test, practice

MOTS CLÉS : test diagnostique, test clinique, pratique

## Introduction

Since the fundamental purpose of any diagnostic test is to help determine whether a patient has or does not have a particular condition,<sup>1,2</sup> clinicians should be aware of certain parameters regarding the tests they use if these tests are to be applied most appropriately and effectively in practice. The most basic parameters that need to be established regarding any clinical test are that it demonstrates a sufficient degree of reliability and validity.<sup>3-5</sup> If these two important parameters are not met, then the test's value in assisting clinicians to arrive at a diagnosis, form a treatment plan, or monitor a patient's progress is questionable.<sup>3-5</sup>

Reliability refers to the consistency and repeatability of outcomes as measured by the clinical test.<sup>3,4</sup> This includes an assessment of whether a test result measured by one examiner would also be obtained by a different examiner performing the test on the same subject at the same time (i.e. inter-examiner agreement) or by the same examiner performing the test on the same subject at a different time (i.e. intra-examiner agreement).<sup>3</sup> Validity refers to whether the clinical test is accurate in measuring what it is purporting to measure.<sup>3,4,6</sup> Of the three types of validity, only "criterion" validity is relevant to the evaluation of a clinical test.<sup>6</sup> This involves the comparison of results obtained from the clinical test to those obtained from a

---

\* Assistant Professor and CCRF Research Chair in Neuromusculoskeletal Health. Faculty of Kinesiology and Health Studies, University of Regina. Email: paul.bruno@uregina.ca. Faculty of Kinesiology and Health Studies. University of Regina, Regina, SK S4S 0A2. Canada.  
Tel: (306) 337-3343, fax: (306) 585-4854.

© JCCA 2011

“reference” (i.e. “criterion”) diagnostic test which, although it provides a more accurate assessment of the condition being investigated, is deemed to be too expensive and/or impractical to use routinely in clinical practice. Therefore, most clinical tests are used to classify patients as “positive” or “negative” depending on the presence or absence (respectively) of a particular sign or symptom, which is then presumed to be indicative of the presence or absence of the condition (i.e. a “positive” test result indicates that the patient has the condition). Assessing the validity of a clinical test’s usefulness in this regard requires knowledge of a variety of parameters, all of which are important and must be individually considered by the clinician in order to appropriately interpret the results he/she obtains when performing the test on a patient.<sup>5,7</sup> These parameters include the test’s *sensitivity*, *specificity*, *predictive values*, and *likelihood ratios*.

The *sensitivity* of a clinical test is the proportion of subjects with the condition who are correctly identified by the test and provide a “positive” result.<sup>1,2,6-8</sup> Thus, if the sensitivity is high, a “negative” test result will effectively rule out the condition.<sup>2</sup> The *specificity* is the proportion of subjects without the condition who are correctly identified by the test and provide a “negative” result.<sup>1,2,6-8</sup> Thus, if the specificity is high, a “positive” test result will effectively rule in the condition.<sup>2</sup> The *positive predictive value* is the proportion of subjects with a “positive” test result who are correctly diagnosed, whilst the *negative predictive value* is the proportion of subjects with a “negative” test result who are correctly diagnosed.<sup>1,2,6,8</sup> Since both the condition’s presence (i.e. “present” or “absent”) as well as the test result (i.e. “positive” or “negative”) are categorical in nature, the resulting calculations for these parameters are based on constructing a 2 × 2 contingency table, as illustrated in Figure 1.

There is an important trade-off between these two pairs of parameters. Although the predictive values are more valuable to clinicians since they provide a direct assessment of the usefulness of the test in practice, they are also both influenced by the prevalence of the condition in the population to whom the test is applied.<sup>1,2,6,8</sup> A higher prevalence tends to lead to an increased positive predictive value and a decreased negative predictive value, whilst a lower prevalence tends to lead to an increased negative predictive value and a decreased positive predictive value.<sup>8</sup> Therefore, it is vital that the

		Condition	
		Present	Absent
Test Result	Positive	a	b
	Negative	c	d

a = “True Positives”    Sensitivity =  $a/(a + c)$   
 b = “False Positives”    Specificity =  $d/(b + d)$   
 c = “False Negatives”    Positive predictive value =  $a/(a + b)$   
 d = “True Negatives”    Negative predictive value =  $d/(c + d)$

Figure 1 An illustration of how to calculate the sensitivity, specificity, positive predictive value, and negative predictive value for a clinical test from a 2 × 2 contingency table.

predictive values that are calculated for a clinical test in a particular study sample should not be taken to apply universally. The sensitivity and specificity, on the other hand, are unaffected by the prevalence of the condition, but are not as useful to clinicians since they give little indication as to how good the test is at predicting the correct diagnosis.<sup>1,2,6,8</sup> For these reasons, the use of these four parameters alone can occasionally lead clinicians to make misleading inferences regarding the value of a clinical test and, therefore, the results they obtain when using it in practice.<sup>6</sup>

As a result, two other parameters, namely the *likelihood ratios of a positive and negative test*, have been suggested to be better indicators of the usefulness of a clinical test.<sup>1,2,6</sup> Effectively, these ratios compare the probability of getting a test result if the subject truly had the condition with the corresponding probability if he/she did not. Figure 2 illustrates how to calculate these parameters, as well as describes the general consensus on how to interpret the resulting values.<sup>1,2,6</sup>

Since these ratios effectively summarize the information contained in each of the four previously-described parameters and are not influenced by the prevalence of the condition, they are considered to be more valuable to clinicians.<sup>2,6</sup> In addition, since likelihood ratios (as well as each of the other four parameters) are proportions, they may be expressed as a percentage and should always be presented with an appropriate confidence interval.<sup>1,7,8</sup>

Likelihood ratio of a positive test = (sensitivity)/(1 – specificity)

Likelihood ratio of a negative test = (1 – sensitivity)/(specificity)

- The higher the likelihood ratio of a positive test, the more certain one can be that a positive test result indicates the subject has the condition. A value of 10 or more is considered an indicator that a positive test result is very good at ruling in the condition.
- The lower the likelihood ratio of a negative test, the more certain one can be that a negative test result indicates the subject does not have the disorder. A value of 0.1 or less is considered an indicator that a negative test result is very good at ruling out the condition.
- If a likelihood ratio is close to 1.0, then the test result is not a good indicator whether the subject has (for a positive test result) or does not have (for a negative test result) the condition.

Figure 2 *An illustration of how to calculate the positive and negative likelihood ratios for a clinical test and the general consensus on how to interpret the resulting values.*

### The Prone Hip Extension Test

The Prone Hip Extension (PHE) Test was one of a series of clinical tests developed by Vladimir Janda as a means of evaluating for motor control deficiencies during specific movements which were proposed to be associated with the development of various musculoskeletal pain syndromes.<sup>9,10</sup> Based on his clinical observations, Janda suggested that this particular test could be used as a means of assessing for a particular functional muscle imbalance (variously referred to as “lower crossed syndrome,” “distal crossed syndrome,” or “pelvic crossed syndrome”) that he deemed to be important in the development and/or perpetuation of low back pain (LBP). Despite slight variations in the traditional descriptions of how to perform the PHE Test, the general procedure was to have the patient lie prone and alternately lift each leg away from the table whilst the clinician observes and/or palpates four muscles of interest – namely the ipsilateral gluteus maximus (GM) and hamstring (HAM) muscles as well as the ipsilateral erector spinae (IES) and contralateral erector spinae muscles (CES) – in an attempt to determine their order of activation.<sup>9–12</sup> Although there was some debate as to what the “normal” order of activation should be during the move-

ment, with both the GM and HAM being proposed as the muscle that should become active first, there was general agreement that these two muscles should become active prior to the CES and IES.<sup>11–13</sup> Regardless of this debate, the clinician was instructed to assess whether the erector spinae muscles were readily activated and/or the activation of the GM was delayed, which would be indicative of an “abnormal” motor pattern for this movement.<sup>9–11,13</sup>

As described, the theory behind this “traditional” use of the PHE Test was based primarily on clinical observations. However, the amount of published research supporting and quantifying its clinical usefulness in this regard is sparse. To the author’s knowledge, there are no published studies which have investigated the validity and reliability of determining the motor patterns that patients use during PHE via observation or palpation. Ergo, the accuracy with which clinicians can detect the muscle activation order by either of these methods, as well as whether the results obtained for a particular patient are reproducible by different clinicians, are unknown. It has been shown that both asymptomatic subjects<sup>14</sup> and LBP patients<sup>15</sup> demonstrate a great deal of within-subject variability in the activation orders they use when performing PHE over a series of repetitions (which is how this test is commonly performed in practice), and that the absolute differences in the relative onset times of the four muscles are generally quite small.<sup>15–18</sup> Considering these findings, it seems reasonable to question whether a clinician could actually be expected to accurately detect these small differences in muscle onsets by a method other than electromyography, a concern which has also been expressed by other authors.<sup>16,17</sup>

Even if it was to be demonstrated that the reliability and validity of detecting muscle onsets by observation or palpation were sufficiently acceptable for the test to be used by clinicians in this manner, the actual clinical importance of the activation order a patient uses to achieve PHE is questionable. First, there does not appear to be “normal” nor “abnormal” muscle activation orders for PHE.<sup>14,15</sup> As well, the original contention that a GM onset after that of the erector spinae muscles was “abnormal” also appears to be incorrect as several studies have demonstrated that the GM seems to most commonly be the final muscle to become active during PHE.<sup>14,15,17,18</sup> Indeed, in both asymptomatic subjects and LBP patients, the HAM, IES, and CES appear to generally become active almost simultaneously and in a seemingly random order, followed by the

GM after a delay. Collectively, these points seriously challenge the appropriateness and clinical value of using the PHE Test as it was traditionally described.

More recently, Murphy et al.<sup>5</sup> provided an alternative description of how clinicians should perform and interpret the PHE Test. They proposed that rather than attempting to assess the motor pattern(s) a patient utilizes to achieve the movement, clinicians should instead observe for the presence of the following “abnormal” deviations of the lumbar spine during the movement: rotation of the lumbar spine such that the spinous processes appeared to move toward the side of hip extension; a lateral shift of the lumbar spine toward the side of hip extension; and extension of the lumbar spine. It was suggested that this may be a better indication of suspected “dynamic instability” of the lumbar spine than the traditional use of the test.

Importantly, the inter-examiner reliability of classifying LBP patients as “Positive” and “Negative” based on the presence or absence (respectively) of the three “abnormal” deviations of the lumbar spine motion described above has been found to be good.<sup>5</sup> There is, however, a paucity of published research attempting to explain the underlying motor control strategies that account for the presence or absence of these deviations. A preliminary study using asymptomatic subjects demonstrated that the presence of one or more of the deviations was associated with a significant delay in the onset of the GM.<sup>19</sup> Although it is unknown at present if these findings are generalizable to the LBP population, they would *seem* to suggest that the presence of these deviations have the potential to be used as an indirect indicator that an “abnormal” motor pattern is present in the form of a significantly delayed onset of the GM during PHE. However, in order for clinicians to more appropriately interpret their findings when using the PHE Test in this manner, they must first consider the validity of using the presence of these deviations as being “diagnostic” of this underlying motor pattern. To this end, knowledge of the test’s sensitivity, specificity, predictive values, and likelihood ratios is necessary.

#### The Presence of “Abnormal” Lumbar Spine Deviations During the PHE Test as Being Diagnostic of a Significantly Delayed Onset of the GM: A Consideration of its Diagnostic Test Parameters

As outlined in the previous section, it has been shown that

the presence of one or more of the lumbar spine deviations seems to be associated with a significant delay in the onset of the GM during PHE in asymptomatic subjects. It must be stressed, however, that this finding was based on the calculation and comparison of group averages. In other words, the average onset time of the GM during sets of PHE that demonstrated the deviations was compared to the average onset time of the GM during sets of PHE that did not. As an author of this particular paper, I can attest to the fact that not all of the sets classified as “Positive” demonstrated a large delay in GM onset compared to those classified as “Negative.” The same holds true for those classified as “Negative” (i.e. a minority of these sets demonstrated a delay in GM onset comparable to those classified as “Positive”). Thus, the presence of these “false positives” and “false negatives” necessitate the calculation of the test’s sensitivity, specificity, predictive values, and likelihood ratios in order to comment on the inherent value that “positive” and “negative” test results would have for clinicians. One caveat that must be emphasized at this point is that since only asymptomatic subjects were used in the cited study, values for these parameters cannot be established for the LBP population based on these data. The following calculations are provided merely for illustrative purposes and should not be taken to inform clinical decisions if/when the PHE Test is performed in clinical practice on LBP patients.

It has been highlighted that diagnostic test parameters analyze a test’s ability to diagnose the presence or absence of a condition by the presence or absence of a particular sign or symptom. As such, the first issue that must be addressed is to define the “condition” that the PHE Test is being used to diagnose in this case. Simply stating that the test is attempting to determine the presence or absence of a “significantly delayed GM onset” during PHE is insufficient since the muscle onset time data is continuous in nature and the test parameters require categorical outcomes for the condition (i.e. the condition is either present or absent). It is therefore necessary to select a specific magnitude for the onset delay above which the relative onset of the GM is defined as “significantly delayed” (i.e. the condition is present), and below which it is defined as “not significantly delayed” (i.e. the condition is absent).

Since there is no universally accepted standard that is used to define that a particular muscle’s onset is “significantly delayed” for this movement, the decision as to

Figure 3 The sensitivity, specificity, predictive values, and likelihood ratios of the PHE Test for diagnosing the presence or absence of a “significant delay” in GM onset as calculated from the data collected in Bruno et al. (2008).<sup>19</sup>

		“Significantly Delayed” GM Onset	
		Present	Absent
PHE Test Result	Positive	7	6
	Negative	4	26

  

Sensitivity = $7/(7 + 4) = 0.636$	63.6% (95% CI: 35.4 to 84.8)
Specificity = $26/(6 + 26) = 0.813$	81.3% (95% CI: 64.7 to 91.1)
Positive predictive value = $7/(7 + 6) = 0.538$	53.8% (95% CI: 29.1 to 76.8)
Negative predictive value = $26/(4 + 26) = 0.867$	86.7% (95% CI: 70.3 to 94.7)
Likelihood ratio of a positive test = $(0.636)/(1 - 0.813)$	3.394 (95% CI: 1.437 to 7.803)
Likelihood ratio of a negative test = $(1 - 0.636)/(0.813)$	0.448 (95% CI: 0.184 to 0.831)

what magnitude of onset delay to select for the “cut-off” will need to be somewhat subjective for the purpose of this example. Several studies have demonstrated a significant delay in the onset of the transversus abdominis during various arm and leg movements in LBP subjects, with the magnitude of the delay varying from ~60 ms to 165 ms depending on the specific movement the subjects were asked to perform.<sup>20–22</sup> It has been suggested that these onset delays are potentially indicative of motor control deficits that may lead to inefficient lumbar spine stabilization. A delayed onset of 110 ms will therefore be selected as being indicative of a motor control deficit during PHE that represents inefficient lumbar spine stabilization since it is the approximate mid-point of the range provided. Thus, the “condition” will be deemed present (i.e. the GM onset will be deemed “significantly delayed”) if the relative onset delay exceeds 110 ms. Conversely, the “condition” will be deemed absent if the relative onset delay is less than 110 ms.

Using this definition, an analysis of the raw data collected for the cited study (not presented) can be analyzed to categorize the PHE sets classified as “Positive” into “true positives” (i.e. those which demonstrate a “significantly delayed” GM onset) and “false positives” (i.e. those which do not demonstrate a “significantly delayed” GM onset). This reveals 7 “true positives” and 6 “false positives.” Likewise, the “Negative” PHE sets can be categorized into “true negatives” (i.e. those which do

not demonstrate a “significantly delayed” GM onset) and “false negatives” (i.e. those which demonstrate a “significantly delayed” GM onset). This reveals 26 “true negatives” and 4 “false negatives.” These values can then be inserted into a  $2 \times 2$  contingency table and used in the calculations in each of the test parameters (see Figure 3).

These calculated values lend themselves to several implications regarding the appropriateness of using the presence of the previously-described “abnormal” lumbar spine motion patterns during the performance of the PHE Test as being diagnostic of an underlying “significant delay” in GM onset. Although the specificity indicates that the value of a positive test result at ruling in the condition is relatively high, only 53.8% of the positive test results were correct. In addition, whilst 86.7% of the negative test results were correct, the sensitivity indicates that the value of a negative test result at ruling out the condition is only moderate. Since predictive values are affected by the prevalence of the condition, it could be argued that these apparent contradictions may be explained by the relatively low prevalence of the condition in this particular sample of subjects, which may have had an effect to raise the negative predictive value and lower the positive predictive value. This reinforces the need to consider the calculated likelihood ratios, both of which indicate that the value of positive and negative test results are “somewhat useful.”<sup>6</sup>

It would be essential for clinicians to know and con-

sider all of these factors if they decide to use the PHE Test in this manner since they would be able to more appropriately interpret and use the test results obtained for a particular patient in their overall clinical decision making. It should be emphasized once again, however, that the implications described above should not be generalized to the LBP population since the parameters have only been calculated here for illustrative purposes.

### Conclusion

It is essential that, whenever possible, clinicians know and make use of specific parameters related to the diagnostic tests they use in practice in order to appropriately interpret their clinical findings. This would allow for more informed decision making when it comes to the diagnosis and management of a patient's condition. Although these parameters have been established for many tests used in health care, this is not universally true. Henceforth, it is crucial that further research be conducted to establish these parameters for the variety of clinical tests that are used by health care practitioners.

### Acknowledgements

The author would like to acknowledge Jeff Bagust, Jennifer Bolton, Jonathan Cook, and Neil Osborne at the Anglo-European College of Chiropractic for the assistance they provided in the conduction of the research that formed the framework for this paper.

### Declaration

The research that formed the framework for this paper was conducted whilst the author was a Research Fellow at the Anglo-European College of Chiropractic. The funding for this fellowship position was provided by the Anglo-European College of Chiropractic and The College of Chiropractors. The author has no conflicts of interest to declare regarding this paper or the material described therein.

### References

- 1 Altman DG. Diagnostic Tests. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence* 2nd Ed. Bristol: BMJ Books; 2000. p. 105–19.
- 2 Davidson M. The interpretation of diagnostic test: a primer for physiotherapists. *Aust J Physiother.* 2002; 48(3):227–32.
- 3 Khan KS, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG.* 2001 Jun; 108(6):562–7.

- 4 Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, et al. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. *Spine.* 2004 Oct 1; 29(19):E413–25.
- 5 Murphy DR, Byfield D, McCarthy P, Humphreys K, Gregory AA, Rochon R. Interexaminer reliability of the hip extension test for suspected impaired motor control of the lumbar spine. *J Manipulative Physiol Ther.* 2006 Jun; 29(5):374–7.
- 6 Chien PF, Khan KS. Evaluation of a clinical test. II: Assessment of validity. *BJOG.* 2001 Jun; 108(6):568–72.
- 7 Bland M. *An Introduction to Medical Statistics.* Oxford: Oxford University Press; 1987.
- 8 Altman DG. *Practical Statistics for Medical Research.* London: Chapman & Hall; 1991.
- 9 Jull GA, Janda V. Muscles and Motor Control in Low Back Pain: Assessment and Management. In: Twomey LT, Taylor JR, editors. *Physical Therapy of the Low Back.* New York: Churchill Livingstone; 1987. p. 253–78.
- 10 Janda V. Evaluation of Muscular Imbalance. In: Liebensohn C, editor. *Rehabilitation of the Spine: A Practitioner's Manual.* Baltimore: Lippincott Williams & Wilkins; 1996. p. 97–112.
- 11 Lewit K. *Manipulative Therapy in Rehabilitation of the Locomotor System.* 2nd ed. Oxford: Butterworth-Heinemann; 1991.
- 12 Chaitow L, DeLany JW. *Clinical Application of Neuromuscular Techniques. Volume 2 – The Lower Body.* Edinburgh: Churchill Livingstone; 2002.
- 13 Janda V, Frank C, Liebensohn C. Evaluation of Muscular Imbalance. In: Liebensohn C, editor. *Rehabilitation of the Spine: A Practitioner's Manual 2nd Ed.* Baltimore: Lippincott Williams & Wilkins; 2007. p. 203–25.
- 14 Bruno P, Bagust J. An investigation into the within-subject and between-subject consistency of motor patterns used during prone hip extension in subjects without low back pain. *Clin Chiropr.* 2006; 9(1):11–20.
- 15 Bruno P, Bagust J. An investigation into motor pattern differences used during prone hip extension between subjects with and without low back pain. *Clin Chiropr.* 2007; 10(2):68–80.
- 16 Pierce MN, Lee WA. Muscle firing order during active prone hip extension. *J Orthop Sports Phys Ther.* 1990; 12(1):2–9.
- 17 Lehman GJ, Lennon D, Tresidder B, Rayfield B, Poschar M. Muscle recruitment patterns during the prone leg extension. *BMC Musculoskelet Disord.* 2004 Feb 10; 5:3.
- 18 Sakamoto AC, Teixeira-Salmela LF, de Paula-Goulart FR, de Moraes Faria CD, Guimaraes CQ. Muscular activation patterns during active prone hip extension exercises. *J Electromyogr Kinesiol.* 2009 Feb; 19(1):105–12.
- 19 Bruno P, Bagust J, Cook J, Osborne N. An investigation into the activation patterns of back and hip muscles during prone hip extension in non-low back pain subjects: Normal

- vs. abnormal lumbar spine motion patterns. *Clin Chiropr.* 2008; 11(1):4–14.
- 20 Hodges PW, Richardson CA. Inefficient muscular stabilization of the lumbar spine associated with low back pain. A motor control evaluation of transversus abdominis. *Spine.* 1996 Nov 15; 21(22):2640–50.
- 21 Hodges PW, Richardson CA. Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *J Spinal Disord.* 1998 Feb; 11(1):46–56.
- 22 Hodges PW, Richardson CA. Altered trunk muscle recruitment in people with low back pain with upper limb movement at different speeds. *Arch Phys Med Rehabil.* 1999 Sep; 80(9):1005–12

## Canadian Chiropractic Research Foundation



Creating a culture of research