Editorial

Towards Better Evaluation Tools

Raymond Chong*

Department of Physical Therapy, Georgia Health Sciences University, Augusta, Georgia, USA

Editorial

Clinicians practice their trade by using a set of diagnostic tests to evaluate and quantify a patient's impairments. In the early days, a good number of these tests were developed from observations and patient interactions. At least that's how many of them started before being refined and implemented. Initially, while the test may have served its purpose, it is not unusual for its limitations to become apparent over the years. In this editorial, I discuss what researchers may need to resort to if they encounter such a test.

A test may appear to have limited functionality if it was developed from a group of patients with the same diagnosis and is then used to evaluate other types of patients. This may happen when no suitable alternative is available and the test is adopted or modified for the sake of expediency as well as on the basis of its face validity. The (standing) functional reach test is one such example. It was developed to assess balance control in the elderly population [1] and was subsequently found to correlate with frailty and history of falls. Although its utility in measuring stability limits has come into question [2], other researchers have determined the test to be applicable in predicting falls in the hemiplegic population [3] but not in people with Parkinson's disease [4].

Another reason why a seemingly popular test may reveal its limitations is due to its lack of ecological validity. This occurs when the test that was developed in the lab fails in its supposition that it is indicative of a person's well-being.

A new test may also be developed because it is easier to implement, can be administered faster, and/or is cheaper. What about accuracy? It is a very important factor to consider. It does not make sense to develop a new test if it is not at least comparable to the gold standard in making a diagnosis or prognosis.

How then does one assess the utility of the new test? It is quite common to evaluate a new test by comparing it to the industry gold standard. Ideally, the new test and the gold standard are both used either simultaneously or within the same test session to evaluate the patient. The extent to which the new test matches with the gold standard's appraisal of the patient determines its suitability in replacing the gold standard. A dilemma is created however, when the validity of the gold standard itself is questionable due to the limitations described above. A recent example of such a predicament is instructive: In Parkinson's disease (PD), patients develop postural instability as the condition progresses, resulting in poor control of their balance in standing and walking activities. At this point, the patient is rated at stage 2.5 or 3 of the disease condition. The gold standard for assessing postural instability in PD is to administer the Pull test in which the patient is pulled backwards forcefully at the shoulders to induce a stepping reaction. The test is negative if the patient takes one or two backward steps and recovers unaided. The test is positive if the patient takes multiple steps (retropulsion) and/or has to be caught by the tester to stop an impending fall. Over the years, researchers and clinicians have come to recognize that the test does not seem to correlate well with the patient's postural control [5-8]. It is not unusual for the Pull test to be negative when the patient exhibits postural instability in the course of interacting with the clinician.

An attempt was therefore made to address the limitation of the Pull test by the creation of a three-part questionnaire [9,10]. In order for the questionnaire instrument to assess postural instability, it was necessary to first establish cut-off scores for the instrument. Since the researchers were also interested in comparing the instrument to the Pull test, this necessitated using the Pull test to classify the postural status of the patients. You can see where this problem was headed. As mentioned earlier, since patients often report experiencing instability even though the Pull test is negative, this methodology resulted in a number of false-negative subjects being placed in the stable group. Thus, the new instrument would likely have registered higher sensitivity and specificity values if not for the apparently low sensitivity and specificity of the Pull test. The potentially high diagnostic value of the instrument was weakened by the use of the Pull test in grouping the PD subjects to develop the instrument.

One way to get around a problematic gold standard when evaluating a new test is to do away with comparing the new test to the gold standard. Instead, one may have to rely on advancements in technology that give rise to more refined and deterministic information to develop the new test. One can also take advantage of the massive number of published research studies to develop a new test. This was how the contents of the three-part questionnaire came about.

Researchers and clinicians must be willing to develop new evaluation tools when the limitations of the original instrument become apparent rather than perpetuate its utility out of respect for the developers of the instrument, tradition, or worse, non-critical application. This may be easier said than done. The challenge to discard the gold standard can complicate grant and manuscript peer review processes when the gold standard, despite its limitations, is endorsed by panel experts and becomes part of a recommended set of evaluation [11].

Open-access journals may be one of the vehicles by which such difficulties may be alleviated. As new knowledge is disseminated quickly and made freely accessible in such a publication model, there is a better chance of having new knowledge becoming incorporated into best-practice recommendations by experts in their field of research or clinical practice.

References

1. Duncan PW, Weiner DK, Chandler J, Studenski S (1990) Functional reach: a new clinical measure of balance. J Gerontol 45: M192-M197.

***Corresponding author:** Raymond Chong, Department of Physical Therapy, Georgia Health Sciences University, Augusta, Georgia, USA, Tel: 706-721-2141; Fax: 706-721-3209; E-mail: rchong8@hotmail.com

2. Jonsson E, Henriksson M, Hirschfeld H (2003) Does the functional reach test reflect stability limits in elderly people? J Rehabil Med 35: 26-30.

3. Takatori K, Okada Y, Shomoto K, Shimada T (2009) Does assessing error in perceiving postural limits by testing functional reach predict likelihood of falls in hospitalized stroke patients? Clin Rehabil 23: 568-575.

4. Behrman AL, Light KE, Flynn SM, Thigpen MT (2002) Is the functional reach test useful for identifying falls risk among individuals with Parkinson's disease? Arch Phys Med Rehabil 83: 538-542.

5. Bloem BR, Beckley DJ, van Hilten BJ, Roos RA (1998) Clinimetrics of postural instability in Parkinson's disease. J Neurol 245: 669-673.

6. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. (2003) Mov Disord 18: 738-750.

7. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, et al. (2004) Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord 19: 1020-1028.

8. Jacobs JV, Horak FB, Van Tran K, Nutt JG (2006) An alternative clinical postural stability test for patients with Parkinson's disease. J Neurol 253: 1404-1413.

9. Chong RK, Morgan J, Mehta SH, Pawlikowska I, Hall P, et al. (2011) Rapid assessment of postural instability in Parkinson's disease (RAPID): A pilot study. Eur J Neurol 18: 260-265.

10. Chong RK, Lee KH, Morgan J, Mehta SH, Hall P, et al. (2012) Diagnostic value of the rapid assessment of postural instability in Parkinson's disease (RAPID) questionnaire. Int J Clin Pract 66: 718-721.

11. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, et al. (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord 23: 2129-2170.