Standardizingsciatic function index rating protocol improves inter-rater reliability and precisionof novice assessors

Hend Wageh[1],Tasneem Mohammad[1] , Mirna M. Elias[1] ,Nagwa Abu Elwafa[1] , Mahmoud Nabhan[1],Nabila S. Hassan [2],Aliaa Rehan Youssef [1]

[1]Department of Physical Therapy for Musculoskeletal Disorders and Surgery/ Faculty of Physical Therapy/ Cairo University[2]Pathology Department/ National Research Center

Abstract

Background: Sciatic function index (SFI) is a simple kinematic analysis used to quantify animals' functional recovery following sciatic nerve and related musclesinjuries. However, the reliability of this assessment method may vary based on experience level and practice. Therefore, this study investigated whether standardization of rating would improve interrater reliability and precision among novice assessors with no previous experience.

Methods: Standardization of SFI measurement was based on the delphi technique. A pilot measurement session was done by four raters (one expert and three novices) on 24 traces, not included in the actual analysis. Then a standardized protocol was developed and agreed upon by consensus through discussion and voting. The four raters then independently assessed 20 other traces obtained from 20 rats; 10normal and 10 animals with a unilateral tibialis anterior induced strain injury.

Results: There was no significant differences between the SFI scores of the expert and each of the novice examiners (P-value > 0.05).Furthermore, there was a strong significant positive correlation between expert and novice assessors (r >0.80, P value< 0.001). Yet, the limit of precision was wide and ranged between26.7 to 29.4%.

Conclusion: Standardized SFI rating protocol increased the inter-reliability of novice compared to expert raters; however, precision is still low and need further improvement.

Keywords: Sciatic function index; Inter-rater reliability, Standardization

Introduction

Sciatic functional index (SFI) is a simple kinematic analysis that assessesthe return of rodents' gait to normal followingsciatic nerve and related muscles injury[1-5]. SFI has the advantage of being simple, easy to administrate and economic[1].This performance test is dependent on the clarity of animals' feet print and, hence, the ability of the assessor to select the region of interest for measurement, thus, assessor's experience may play a role in measurement accuracy and precision[6].

SFI intra-rater reliability was found high (r >0.80)between novice and expert raters using semi-automated instruments, where raters were requested to only identify reference points before the software automatically do all the required calculations [6], yet the reliability testing was done on animals with sciatic nerve injury. To authors' knowledge, inter-rater reliability was not done for manual measurement where the assessors select the reference points and do the measurement using the free hand technique of the software. Further, its inter-reliability to assess functional recovery following muscle injury has never been established. Muscle injury is different from that of nerve as it iscommonly associated with pain that may persist for a prolonged period of time, interfering with animal's ability to bear weighton the injured limb. This in turnsmay alter the clarity of feet print in various forms. Distorted feet print may be difficult to identify by novice assessors with no previous experience, resulting in high errors. A standardized assessment protocol may improve noviceassessors' precision and reduce measurement error. Thus,this study aimed atinvestigating whether a standardized measurement protocol would improve interrater reliability, agreement and precision of SFI measurement between expert and novice raters using manual measurements.

Material and methods

Study design

Observational study on a murine animal model

Participants

SFI measurement was done by four assessors, an expert assessor with 3 years of experience in SFI measurement and three novice assessors who have never measured the SFI before.

Measurement procedures

SFI traces were collected from 20 adult male Wister rats; 10 normal and 10 rats that underwent a unilateral strain of tibialis anterior muscle using a valid non-invasive strain induction protocol[4]**.** Strain induction and SFI recording were done in accordance and approved bythe Cairo University Institutional Animal Care and Use Committee (IACUC) guidelines.

SFI recording:

To record SFI, one meter wooden walkway waslined with a squared paper sheet that was fixed in place using adhesive tape. An animal cage was placed at the end of the track to motivate the animal to walk in this direction. Then,animal's paws were soaked in a blue ink before animalswere left to walk across the walkway[4]. Animal conditioning on testing was done before they were allowedto walk for 1-5trials;until the most representative clear feet printswere obtained. All trials were recorded simultaneously by a video camera to confirm walking print selection for further analyses. Distorted prints were discarded, for example, if they were stationary or unclear due to tail drag, smeared, or overlapped with front limbs traces[7, 8]**.**

Standardization of measurement was based on available literatures and was employed using the delphi technique[9]. Briefly, a pilot measurement session was done on 24 traces, not included in the analysis, by the four assessors. Then, all disagreements were extensively discussed and voting on standardization of reference points, especially distorted prints and artifacts, was done. For example, if the heel print was distorted, then the clearest distal heel central point was taken as a reference point**(Fig. 1b-c).** Further, toenail marks or ink spread, especially at the heel, were excluded [7]**.** This was also true for unclear prints.

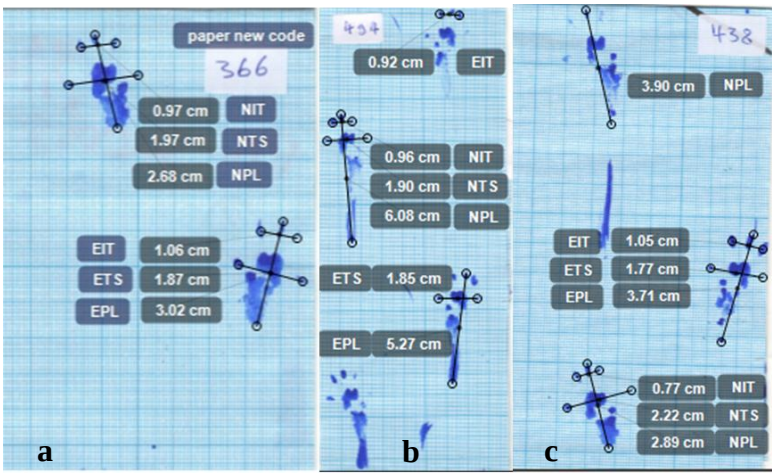

Figure 1: Measurement of different variables of SFI shown on traces obtained from three different rats. E is the experimental side and N is the normal side, **a:**showing a clear PL print that was measured to the edge of the center of the heel, **b:**A distorted PL trace that was measured from the tip of 3rd toe to the most distal point, **c:**A distorted print in which the line between experimental and normal paws was excluded due to inconsistencies.

After the development of a standardized protocol, each assessor selected 3-6 traces from all collected trials of each animal to measure.Then, selected traces were masked and scanned by an investigator that was not involved in rating. Scanned prints were saved as images on personal computer and were then measured at a randomorderusing the Kinovea software; an open access application for the movement evaluation and analysis[10]**.**

The SFI was calculated according to Bain'sequation [8],where E represents the experimental side and N represents the normal side as follows:

SFI = -38.3 × (EPL – NPL/NPL) + 109.5 (ETS-NTS/NTS) + 13.3 (EIT – NIT/NIT) – 8.8

The SFI possible scores ranges from 0-100%, with 0 indicating normal function and 100% indicating complete loss of function.Threevariables were measured for each print as follows [1]: (1)**Print length (PL):**the distance between the tip of the third toe and the most posterior part of the foot in contact with the ground, (2) **Total Toe Spreading (TS):** the distance from the center of the first toe to the center of the fifth toe, and (3) **Intermediary toes (IT):**the distance from the center of the second toe to the center of the fourth toe. All variables of interest were measured to the nearest millimeter.The largest value for each variable was used to calculate theSFI[1].All prints were measured independently by the four assessors and data was recorded in excel worksheets.

Data analysis

The primary outcome for this study was the inter-rater reliability between the expert and novice assessors in measuring SFI. Non-parametricWilcoxon signed rank test was used to investigate the difference between expert and novice examiner's values. Further, Spearman's correlation coefficient was done to detect the magnitude, strength and direction of the association between the results of expert and novice examiners. Finally, Bland-Altman graphical method was plotted to assess the agreement between examinersas well as the degree of precision[11].

Results

Expert SFI scores were not significantly different from each of the novice raters(table 1,P-value >0.05). Further, a significantly strongpositive correlation between expert and each of the novice raters' scores was found (r>0.8, P-value <0.001).

The limit of agreement (2SD) between expert and novice examiners was 95%,whereas the limit of precision (1SD) ranged between 90 and 95%.The precision width ranged between 26.7 and 29.4%**(Figures 2 through4)**.

Table 1: SFI mean and standard deviation for each observer for each group

Examiners	Groups' SFI mean and standard deviation	
	Normal control	TA injured group
Expert	-9.37 ± 24.45	-11.33 ± 11.72
Novice rater 1	-13.85 ± 15.79	-10.31 ± 8.28
Novice rater 2	-16.1 ± 13.12	-13.02 ± 8.24

Novice rater 3	-16.69 ± 14.21	-12.81 ± 7.97

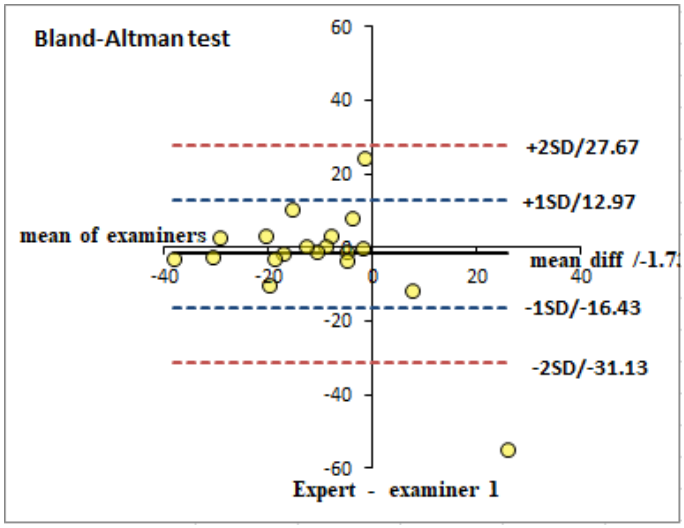

Figure 2: Bland-Altman agreement plot between the expert and first novice rater

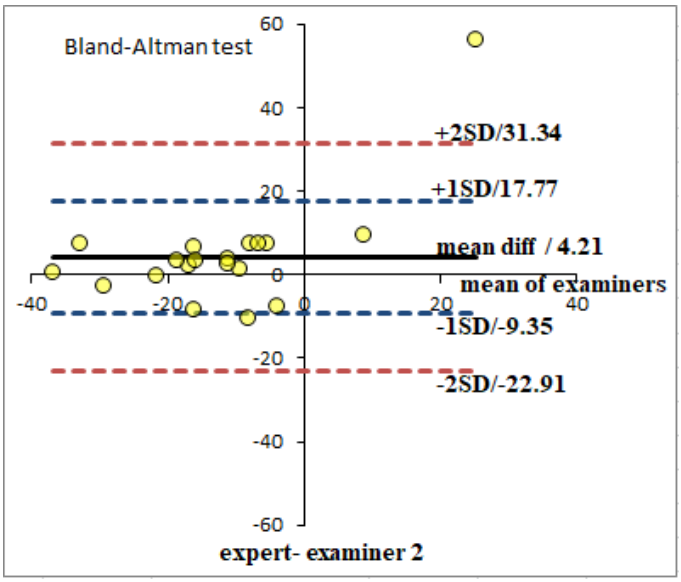

Figure 3: Bland-Altman agreement plotbetween the expert and second novice rater

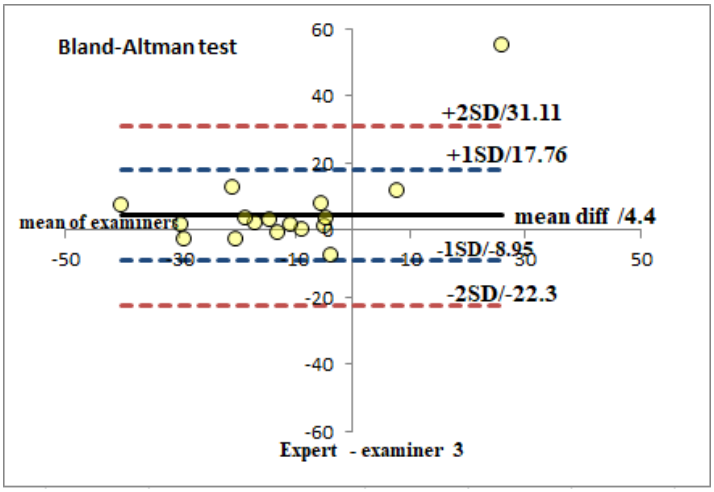

Figure 4: Bland-Altman agreement plotbetween the expert and the third novice rater

Discussion

This study evaluated the efficacy of standardizing SFI measurement on the inter-reliability of novice raters in measuring SFI manually after muscle strain.Results showed no significant differences between the scores of expert and novice raters, denoting improved measurement accuracy in novice raters. This was further confirmed by the

significant positive correlation. Absolute score agreement between examiners was high,which further confirmed the efficacy of the used standardized protocolin reducing the effect of experience on measurements.

However, the degree of precision was low and varied up to almost 30%.This variability is high. Considering that normal SFI value ranges between 0 and 11%[1], then raters may misclassify functional recovery,as normal or abnormal, with this wide variability in measurement.It should be emphasized that SFI normal range was identified based on other calculation equation[1] that was not used in this study, thus normal ranges should be confirmed for current equation.

For reliability, the result of the correlation reported in this study is comparable to that reported by Monte-Raso et al. (2008). In their study, four assessors, three novice and one expert, measured SFI following sciatic nerve injury. A strong positive correlation (r=0.82)was reported at the 3rd week following injury, whereas the first two weeks after injury had a very low correlation due to the poor quality of feet prints.However, Monte-Raso and his colleagues obtained rats' feet print on different type of papers (saturated with a bromophenol blue solution). Further, they used a more advanced, complicated and semiautomated method, so they did not have the problem of ink smearing or distorting the prints and, probably, less measurements subjectivity,For example, SFI calculation was done using a graphic softwarethat required assessors to only identify the reference points but not doing the actual measurements manually.Moreover, Monte-Raso et al. (2008)had a different raters training, mainly on software use, not the whole selection and the manual measurement process that was done in the current study[6].

Brown et al. (1991) also reported an excellentand even greater SFI interrater reliability (r= 0.92) after sciatic nerve surgical repair. Yet, in Brown's study, four experts done the analysis and measurements, whereas in the current study novice raters were included, which are expected to show higher variability associated with the lack of experience. Further, Brown and his colleagues also employed the semi-automated software which, as explained earlier, could reduce the variability associated with manual measurements, yet needs special equipment[12].

To authors' knowledge, this is the first study to investigate SFI inter-rater reliability between novice and expert raters during manual measurement of SFI using an open access software. Further, this study is the first to assess inter-rater reliability of SFI after muscle and not sciatic nerve injury. However, a few limitations exist: First, measurements were entirely dependent upon subjective manual measurement, and not compared to other confirmatory objective measurements such as kinematic and kinetic analyses which may improve precision. Second, animals were given various numbers of trials until the optimum print was collected, repetition of walking trials may have caused animals fatigue, and hence, overestimated the functional impairment, yet this is not expected to affect inter-rater variability.

Based on the results of this study,using a standardized protocolfor SFI manual measurement improved the reliability between assessors, regardless to experience level.However, precision is still low; requiring improved measurement procedures.

References

1. de Medinaceli, L., Freed, W. J., & Wyatt RJ. An Index of the Functional Condition of Rat Sciatic Nerve Based on Measurements made from Walking Tracks. *Exp*

Neurol. 1982;77(3):634-643.

2. Varejão, A. S., Meek, M. F., Ferreira, A. J., Patrício, J. A., & Cabrita AM. Functional evaluation of peripheral nerve regeneration in the rat: walking track analysis. *J Neurosci Methods*. 2001;108(1):1-9.

3. Reynolds, J. L., Urbanchek, M. S., Asato, H., & Kuzon WM. Deletion of Individual Muscles Alters Rat Walking-Track Parameters. *J Reconstr Microsurg*. 1996;12(7):461-466.

4. Ramos, L., Junior, L., Pinto, E. C., Pallotta, R. C., Frigo, L., Marcos, R. L., ... & Lopes Martins RÁB. Infrared (810 nm) low-level laser therapy in experimental model of strain-induced skeletal muscle injury in rats: Effects on functional outcomes. *Photochem Photobiol*. 2012;88(1):154-160.

5. Dos Santos, L. S., Saltorato, J. C., Monte, M. G., Marcos, R. L., Lopes-Martins, R. Á. B., Tomazoni, S. S., ... & de Paiva Carvalho RL. PBMT and topical diclofenac as single and combined treatment on skeletal muscle injury in diabetic rats : effects on biochemical and functional aspects. *Lasers Med Sci*. 2018:1-8.

6. Monte-Raso, V. V., Barbieri, C. H., Mazzer, N., Yamasita, A. C., & Barbieri G. Is the Sciatic Functional Index always reliable and reproducible? *ournal Neurosci methods*. 2008;170(2):255-261.

7. Dellon, E. S., & Dellon AL. Functional assessment of neurologic impairment: track analysis in diabetic and compression neuropathies. *Plast Reconstr Surg*. 1991;88(4):686-694.

8. Bain, J. R., Mackinnon, S. E., & Hunter DA. Functional evaluation of complete sciatic, peroneal, and posterior tibial nerve lesions in the rat. *Plast Reconstr Surg*. 1989;83(1):129-136.

9. Hsu, C. C., & Sandford BA. The Delphi technique: making sense of consensus. *Pract assessment, Res Eval*. 2007;12(10):1-8.

10. Guzmán-Valdivia, C. H., Blanco-Ortega, A., Oliver-Salazar, M. A., & Carrera-Escobedo JL. Therapeutic Motion Analysis of Lower Limbs Using Kinovea. *Int J Soft Comput Eng*. 2013;3(2):359-365.

11. Giavarina D. Understanding bland altman analysis. *Biochem medica*. 2015;25(2):141-151.

12. Brown, C. J., Evans, P. J., Mackinnon, S. E., Bain, J. R., Makino, A. P., Hunter, D. A., & Hare G. Inter- and intraobserver reliability of walking- track analysis used to assess sciatic nerve function in rats. *Microsurgery*. 1991;12(2):76-79.

توحيد التقييملمؤشر عرق النسا الوظيفي يحسن الموثوقيه والدقه للمقيمين المبتدئين

الخلفية: المؤشر الوظيفي لعرق النسا هو تحليل حركي بسيط يستخدم لتحديد التعافي الوظيفيفي الفئران بعد اصابة عصب النسا والعضلات التي يغذيها. ومع ذلك قد تختلف الموثوقية بين المقيمين لطريقة القياس بناء علي مستوي الخبره والممارسة. ولذلك ، هدفت هذه الدراسة اليالتحقق منانه إذا تم توحيد التقييمسينعكس هذا علي تحسن الموثوقية والدقه بين المقيمين المبتدئين الذين ليس لديهم خبرة سابقة في هذا القياس.

الطريقة:تم توحيد طريقة قياس المؤشر الوظيفي لعرق النسا بطريقة ديلفي بناء علي قياس اثار ٢٤ فأر منخلال تجربة استطلاعية ثم قام المقيمون الأربعة بشكل مستقل بتقييم آثار عشرون فأر ؛ عشرة طبيعين و عشرة حيوانات تم استحداث اصابة التمزق العضلي لعضلة الظنبوبي الأماميه في جانب واحد لهم.

النتائج: لم يكن هناك فروق ذات دلاله احصائيه بين درجات المؤشر الوظيفي لعرق النسا للخبير وكل من الفاحصين المبتدئين (مستوي الدلالة > ٠.٠٥). علاوة علي ذلك ، كان هناك ترابط ايجابي قوي ذات دلاله احصائيه بين المقيم الخبير والمبتدئين (معامل الأرتباط > ٠.٨، مستوي الدلالة < ٠.٠٠١). ولكنكانت حدود الدقة واسعة وتراوحت بين ٢٦.٧ إلى ٢٩.٤٪.

الخلاصة:توحيد طريقة قياس المؤشر الوظيفي لعرق النسا أديالي زيادة موثوقية المقيمين المبتدئين؛ ولكن، الدقة لا تزال تحتاج الي مزيد من التحسين.

الكلامات الدالة: المؤشر الوظيفي لعرق النسا ؛ الموثوقية بين المقيمين ؛ التوحيد.