SHORT COMMUNICATION

HAND-HELD DYNAMOMETRY: TESTER STRENGTH IS PARAMOUNT, EVEN IN FRAIL POPULATIONS

Carol A. Stone, MSc[1], Bríd Nolan, MSc[1], Peter G. Lawlor, MB, MMedSc[2] and
Rose Anne Kenny, MD[3]

From the [1]Our Lady's Hospice & Care Services, Harold's Cross, Dublin, Ireland, [2]Division of Palliative Care, University of Ottawa & Palliative Care Unit, Bruyere Continuing Care, Canada and [3]Department of Medical Gerontology, Trinity College Dublin & Centre of Excellence for Successful Ageing, St James' Hospital, Dublin, Ireland

Objective: To determine test-retest and inter-rater reliability of hand-held dynamometry when used to measure knee-extensor strength in patients with advanced cancer.

Subjects: Adults with metastatic or locally advanced cancer recruited from palliative care services to a study of the risk factors for falls.

Methods: Consecutive recruits ($n=30$) underwent repeat testing after an interval of 1 h, by the same researcher, to assess test-retest reliability. The subsequent 15 patients underwent retesting by a second researcher. The intra-class correlation coefficient and limits of agreement were calculated.

Results: The test-retest reliability difference between measurements increased with the magnitude of measurement, mean leg strength=113 N (standard deviation 43.1), 95% ratio limits of agreement 0.81–1.5, intra-class correlation coefficient=0.9. The inter-rater testing mean leg strength=128.5 N (standard deviation 35.1), 95% limits of agreement=–57.24 to 36.06 N. Intra-class correlation coefficient=0.83.

Conclusion: Test-retesting and inter-rater testing yielded high intra-class correlation coefficients, but the limits of agreement were wide. In test-retesting, the difference between tests increased as the magnitude of measurement increased. It has been widely reported that hand-held dynamometry is reliable when used to measure knee-extensor strength in frail or elderly persons. However, our results show that, even in these populations, reliability may be compromised by inadequate tester strength.

Key words: research methods; dynamometry; muscle strength; reliability; validation; neoplasm.

INTRODUCTION

Loss of skeletal muscle mass is a recognized feature of ageing, and lower limb muscle weakness a recognized risk factor for falls in older persons (1, 2). In advanced cancer, muscle strength may be adversely affected by many factors, including anorexia-cachexia-associated skeletal muscle wasting, immobilization and proximal myopathy related to cortico-steroid treatment. We conducted a prospective study of the risk factors for falls in patients with advanced cancer. Our selection of independent risk factor variables was informed by a review of the literature on risk factors for falls in older persons. Hence, we wished to include an objective, responsive and reliable measure of lower limb strength as an independent variable. Our selection was also based on the desirability that the method of testing be potentially transferrable to clinical settings, in the event that it was shown to be a risk factor for falls in advanced cancer.

Hand-held dynamometry (HHD) provides an objective measure of muscle strength, and the equipment is generally small and portable. Although demonstratively less reliable when used by testers of below-average strength to test large muscle groups (such as the knee extensors of young healthy subjects), it is widely reported to be reliable in testing muscle strength of older or infirm individuals (3, 4). It has been shown to have good test-retest reliability when used to measure knee extensor strength in 10 community-dwelling elderly persons (5), 41 community-dwelling older persons with a history of falling (6), and 13 patients referred for domiciliary physiotherapy (7). Bohannon & Andrews (8) also reported a high level of inter-rater reliability of HHD when used to measure the strength of knee extensors in mostly post-stroke patients undergoing physiotherapy.

Based on our expectation that knee extensor strength, and hence reliability, of HHD would be similar in our patient cohort to that of older or frail persons, we elected to measure knee extensor strength using HHD. The testing protocol, including the positioning of transducer and subject, were informed by trials with healthy subjects. We report here the results of reliability testing, the aim of which was to establish inter-rater and test-re-test reliability of measurements taken by the two testers involved in the research project.

METHODS

Patients aged over 18 years with a diagnosis of cancer that is metastatic or locally advanced, admitted consecutively to home-care, day-care and inpatient palliative care services, were screened for eligibility for inclusion in the study of the risk factors for falls. Exclusion criteria included: being unable to sit-to-stand or to mobilize independently

for a distance of 6 m, or being considered too unwell to participate, or actively dying by the admitting physician and research team. The study was approved by St Vincent's University Health Group ethics committee.

During the validation phase of the project, consecutive patients underwent repeat testing of right (R) leg strength. The testing protocol was as follows: the subject was seated, hips and knees at 90°, hands resting on the tops of their thighs. Following verbal explanation, the dynamometer was placed 10 cm distal to the tibial tuberosity and the subject asked to "straighten your leg as strongly as you can, stronger, stronger, release" (4 s). The MicroFET 2 (Hoggan Health Industries, West Jordan, Utah, USA) was used. The maximal force was noted and the best of 3 measurements was recorded as the result. Both tester and subject were blinded to the result. The test was repeated 1 h later, by the same tester for the first 30 subjects and by the second tester for the next 15 subjects, with the testers alternating as to who tested first.

Statistical analysis

Agreement between the two measurements was examined by calculation of the limits of agreement from the mean difference/bias (D) and the standard deviation (SD) of the differences. The interpretation is that, for a new individual from the studied population, there is 95% probability that the difference between any two measurements should lie within the limits of agreement (9–11). The data was first checked for heteroscedasticity (whether the differences depend on the magnitude of the measurement) by examination of mean-vs-differences plots and calculation of the corresponding Kendall's correlation coefficient. In the event that heteroscedasticity was present, the data underwent logarithmic transformation and reassessment for resolution of the relationship between log difference and log mean and the geometric standard deviation was calculated. In this case, the 95% ratio limits of agreement were calculated by division and multiplication of the mean difference by the square of the geometric standard deviation (GSD), the interpretation being slightly different, as described below (11, 12). Intra-class correlation coefficients (ICC) were calculated (13).

RESULTS

Test-retest reliability

The mean age of participants was 60 years (SD 12.5), and 18/30 were male. Mean right leg strength was 113 N (SD 43.1) (equivalent to 11.5 kg). Fig. 1 shows the mean-vs-absolute differences plot: Kendall's tau = 0.33, $p = 0.01$. Fig. 2 shows the mean-vs-absolute differences plot for transformed
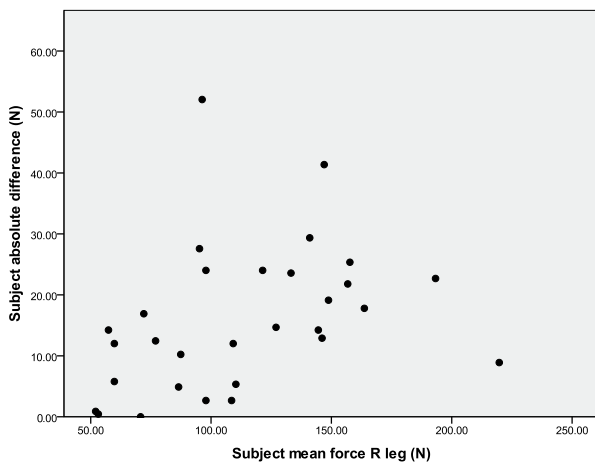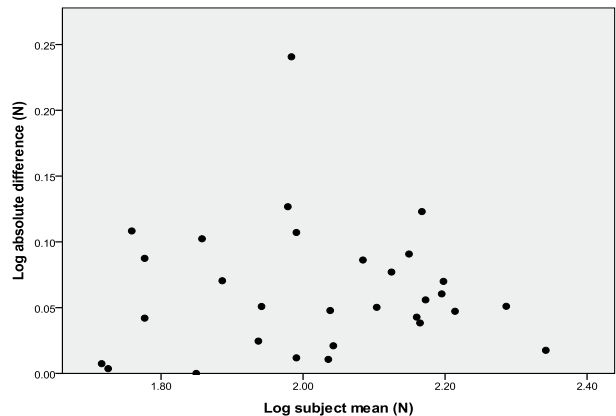

Fig. 2. Test-retest reliability; mean-vs-differences plot after log transformation.

data: Kendall's tau = 0.03, $p = 0.84$. Mean difference = 1.1, $GSD^2 = 1.36$, ratio limits of agreement are 0.81–1.5 (1.1 ÷ 1.36 to 1.1 × 1.36). ICC = 0.9.

Inter-rater reliability

The mean age of participants was 69 years (SD 9.6), and 7/15 were male. The mean right leg strength was 128.5 N (SD 35.1) (equivalent to 13.1 kg). See mean-vs-absolute differences plot (Fig. 3): Kendall's tau = 0.18, $p = 0.35$. D = −10.59 N (SD = 23.8), 95% limits = −10.59 N ± 46.65 = −57.24 to 36.06 N. ICC = 0.83.

DISCUSSION

Analysis of test-retest data and inter-rater data yielded ICCs of 0.9 and 0.83, respectively. The ICC provides an estimate of the proportion of the total variance that is accounted for by the variation between subjects; the remaining variance being attributable to the variation between repeated measurements within subjects (13). The ICC alone provides useful, but incomplete, information regarding reliability, as it gives no
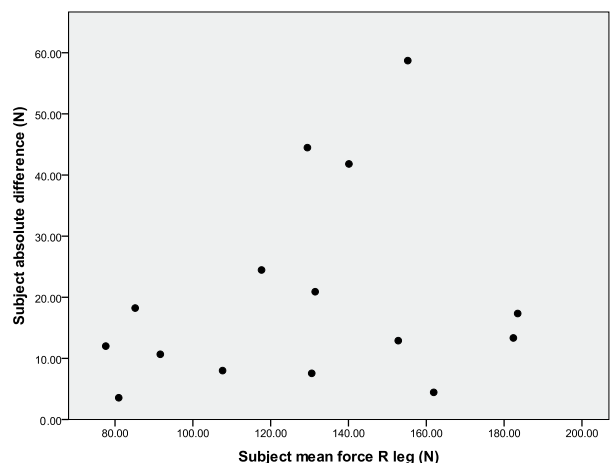

Fig. 1. Test-retest reliability; mean-vs-differences plot (*n* = 30).

Fig. 3. Inter-rater reliability; mean-vs-differences plot (*n* = 15).

sense of the actual magnitude of within-subject differences. An additional drawback is that its value is influenced by the variance in the sample population.

The second statistical method that we used produces an absolute measure of reliability; the limits of agreement provide an estimate of the 95% confidence intervals for the mean difference or bias between two measurements, assuming that the difference is constant and does not vary with the size of the measurement. If logarithmic transformation is required in order to satisfy this assumption, 95% ratio limits of agreement are generated.

The statistical measures, used by Schaubert & Bohannon (5) and Bohannon (7) to express absolute reliability were the coefficient of variation (CV) and the technical error of the measurement (TEM). The former equates to 100×[within-subject SD/ sample mean] and the latter approximates the within-subject SD. Atkinson & Nevill (11) argue that reliability measures based on 1 SD are inadequately useful and that, instead, in the case of CV, the sample SD should be multiplied first by 1.96 before being expressed as the CV, in order to cover 95% rather than 68% of the repeated measures. For the Schaubert & Bohannon (5) and Bohannon (7) data this would yield CVs of approximately 22.8% and up to 25%, respectively, from which one may draw less confident conclusions regarding the test-retest reliability of HHD to measure leg strength in older or frail persons.

The 95% ratio limits of agreement for the test-retest data were 0.81–1.5; hence, for any individual within the population, there is a 95% probability that any two tests will differ due to measurement error by no more than 19% in a negative direction or 50% in a positive direction. Analysis of the inter-rater reliability data shows that tester one's measurements were on average 10.59 N less than those of tester 2 and that for measurements taken on a new subject within the target population, there is 95% probability that the difference between the two testers would be between −57.24 and 36.06 N.

To summarize, test-retesting and inter-rater testing of HHD for measurement of knee extensor strength in patients with advanced cancer yielded high ICCs, but the limits of agreement were wide relative to the mean measurement. Inspection and analysis of the test-retest data revealed increasing difference between tests as the magnitude of measurement increased, suggesting that our less than satisfactory results were at least in part due to stronger subjects' ability to overcome tester strength. Whilst it has been widely written that HHD is reliable when used to test muscle strength in frail or elderly populations, it is clear that tester strength is as important a determinant of reliability as the characteristics of the sample being tested. The mean knee extensor strength of community-dwelling elderly fallers tested by Wang et al. (14) using very similar methods was comparable to that of our own sample. In contrast to our results however, test-retesting of the right leg yielded an ICC of 0.99 and limits of agreement of ± 14.8 N (standard error of the mean × 1.96 × √2) (14). In order to investigate the effect of tester strength on test-retest and inter-rater reliability, Wilholm & Bohannon (3) used 3 testers with measurably different strengths to measure strength in 3 muscle groups in

27 healthy adults. They found HHD testing for muscle groups with a mean force of up to 120 N to be reliable regardless of tester strength (3). This is equivalent to the mean strength of knee extensor measurements in our sample, but despite our having taken the step of placing the transducer more proximal to the knee than typically described, in order to maximize the lever arm to give best mechanical advantage to the tester, we were unable to demonstrate adequate test-retest or inter-rater reliability. Patient characteristics may also have negatively impacted on our results: in advanced cancer, fatigue characterized by reduced endurance and abnormal muscle metabolism is common and may have impacted upon participants' ability to make a consistent maximal effort (15). Alternatively, the consistency of effort may have been negatively affected by discomfort at the site of transducer placement, mentioned by some of the participants in this study and also noted by Kelln et al. (4).

In conclusion, published results of reliability testing of HHD to measure muscle strength in frail or older populations are not generalizable, as reliability is significantly influenced by the strength of the tester. In addition, some authors have employed inadequate statistical measures to describe reliability, leading to overly conservative estimates of measurement error. Ideally, medical rehabilitation practitioners or researchers considering using HHD to measure baseline or post-intervention muscle strength, should personally trial the device before purchasing, in order to assess its reliability when used by them to test a sample of their target population. Alternatives, which waive the issue of tester strength, include attachment of the dynamometer to a fixed stable structure or construction of a resistance-enhanced dynamometer. Although neither has the appeal of HHD alone in terms of simplicity and portability, and the latter would require specialist skills, both have been shown to have better test-retest and inter-rater reliability than conventional HHD (16, 17).

ACKNOWLEDGEMENTS

REFERENCES

1. Lord S, Sherrington C, Menz H, Close J. Falls in older people: risk factors and strategies for prevention. 2nd edn. Cambridge: Cambridge University Press; 2007.
2. Moreland JD, Richardson JA, Goldsmith CH, Clase CM. Muscle weakness and falls in older adults: a systematic review and meta-analysis. J Am Geriatr Soc 2004; 52: 1121–1129.
3. Wikholm JB, Bohannon RW. Hand-held dynamometer measurements: tester strength makes a difference. J Orthopaed Sports Phys Ther 1991; 13: 191–198.
4. Kelln BM, McKeon PO, Gontkof LM, Hertel J. Hand-held dynamometry: reliability of lower extremity muscle testing in healthy, physically active young adults. J Sport Rehabil 2008; 17: 160–170.
5. Schaubert KL, Bohannon RW. Reliability and validity of three strength measures obtained from community-dwelling elderly persons. J Strength Condit Res 2005; 19: 717–720.

6. Wang C, Olsen SL, Protas EJ. Test-retest strength reliability: hand-held dynamometry in community-dwelling elderly fallers. Arch Phys Med Rehabil 2002; 83: 811–815.

7. Bohannon RW. Hand-held dynamometer measurements obtained in a home environment are reliable but not correlated strongly with function. Int J Rehabil Res 1996; 19: 345–347.

8. Bohannon RW, Andrews AW. Interrater reliability of hand-held dynamometry. Phys Ther 1987; 67: 931–933.

9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nurs Studies 2010; 47: 931–936.

10. Bailey SM, Sarmandal P, Grant JM. A comparison of 3 methods of assessing inter-observer variation applied to measurement of the symphisis-fundal height. Br J Obstet Gynaecol 1989; 96: 1266–1271.

11. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine.

Sports Med 1998; 26: 217–238.

12. Bland JM, Altman DG. Measurement error proportional to the mean. Br Med J 1996; 313: 106.

13. Peat J, Williams K, Mellis C, Xuan W. Health Science Research; a handbook of quantitative methods. London: Sage; 2002.

14. Bland JM, Altman DG. Measurement error. Br Med J 1996; 312: 1654.

15. Bruera E, Sweeney C. Cachexia and asthenia in cancer patients. Lancet Oncol 2000; 1: 138–147.

16. Fenter PC, Bellew JW, Pitts TA, Kay RE. Reliability of stabilised commercial dynamometers for measuring hip abduction strength: a pilot study. Br J Sports Med 2003; 37: 331–334.

17. Lu T-W, Hsu H-C, Chang L-Y, Chen H-L. Enhancing the examiner's resisting force improves the reliability of manual muscle strength measurements: comparison of a new device with hand-held dynamometry. J Rehabil Med 2007; 39: 679–684.