Assessing lower limb position sense in stroke using the Gradient Discrimination Test (GradDT™) and Step-height Discrimination Test (StepDT™): a reliability and validity study

Abstract

Purpose: To evaluate the psychometric properties of two novel tests of lower limb position sense.

Methods: Our newly developed tests assess discrimination thresholds of under-foot slope and step height perception using a two alternative forced choice approach. Stroke participants (n=32) and age matched controls (n=32) were tested. Inter- and intra-rater reliability and agreement, sensitivity and specificity, discriminant and convergent validity were evaluated.

Results: Intra-rater reliability for both variants of the gradient discrimination test was excellent: intraclass correlation coefficients (ICC) =0.91 and 0.89. The step height discrimination test had excellent intra-rater reliability and agreement: ICC=0.95. Inter-rater reliability was also excellent in both tests (ICC= 0.85-0.93). Discriminant validity was demonstrated with significant differences in test performance between stroke and control participants ($p<0.001$). Our novel tests did not significantly correlate with the proprioceptive component of the Erasmus modified Nottingham Sensory Assessment. Receiver Operating Characteristic curve analysis indicated both novel tests to have greater sensitivity and specificity than the proprioceptive component of the Erasmus modified Nottingham Sensory Assessment in predicting the presence of self-reported sensory impairments. Functional Reach Test, 10 metre walk test, Centre of Pressure measurement and reported falls showed significant and moderate to strong correlations with novel test performance ($r=0.40-0.60$); the Erasmus modified Nottingham Sensory Assessment did not.

Conclusion(s): Our novel, functionally oriented tests of lower limb position sense are reliable, valid and feasible for use in an ambulatory chronic stroke and elderly population.

Key words: stroke, lower-limb, position sense, measurement

Introduction

To respond to the terrain underfoot or clear obstacles, the position of the lower limb must be detected, processed by the Central Nervous System (CNS), and integrated with visual and vestibular inputs [1]. Inputs from muscle, joint, and cutaneous afferent fibres represent the complex somatosensory modality of proprioception, providing a sense of limb movement (kinesthesia), and limb position [2]. Models of motor control [3], studies of motor learning [4] and neuro plastic adaptation [5] implicate the need for accurate proprioceptive information.

Impaired proprioceptive function in the lower limbs is characteristic of several clinical populations and is associated with altered postural control, a varied unstable gait, and increased falls risk [6-8]. Following stroke, 30-56% of people have somatosensory deficits of the lower limbs [9,10], with somatosensory deficits and motor weakness resulting in worse functional outcomes at six months than motor weakness alone [11]. Moreover, difficulty sensing the position of the foot is reported by people with stroke to impact their ability to walk outdoors, maintain balance and is implicated in falls [7,12].

Despite this, evidence from cross-sectional studies of the stroke population is equivocal; several studies highlight only weak associations between lower limb proprioception and functional outcomes [9, 13]. It is postulated that such findings may, in part, be explained by the methods used to quantify this complex sensory modality. Systematic reviews have highlighted multiple measures that attempt to capture either movement or position sense in a variety of clinical and healthy populations [14, 15]. Typically, approaches to measuring proprioceptive ability are manually administered by the clinician, to a sitting or supine patient, and lead to a classification of proprioceptive ability as ‘absent’, ‘impaired’ or ‘normal’. They are subject to unquantifiable and non-standardised movement speeds, varying

tactile input, and questionable accuracy due to the (often visual) estimation of either the extent of limb mismatch error or movement occurring. Methods that use contralateral inter-limb position matching (i.e. compare “affected” limb with “non-affected” limb) may be further confounded by the presence of bilateral somatosensory impairments following unilateral hemispheric stroke [10,16]. Such methods often form the sub-tests of global, multi-modal somatosensory screening measures such as the Rivermead Assessment of Somatosensory Performance [17], Nottingham Sensory Assessment [18,19], and the sensory scale of the Fugl-Meyer Assessment [20], which have proven clinically utility but questionable accuracy [14, 15] and limited responsiveness [21,22]. The upshot is they may be capable of identifying only the most profound proprioceptive deficits and so should be used cautiously [21,23].

More sophisticated measures of lower limb proprioception, which require specialised or automated equipment, have been used in studies of stroke and neurological populations [24-26] with reported advantages [15,26]. However, the use of such measures in clinical environments may be prohibitive due to lack of commercial availability, equipment ease of use, cost and portability [14]. It has thus been suggested that measures which possess greater accuracy tend to lack clinical utility, with the reverse also being true [15]. In summary, there is a lack of proprioception measures that may satisfy both clinical utility and psychometric requirements.

Tests were developed to address the issues raised above: the need for accurate, standardised measures with appropriate psychometric properties that could be used in both clinical and research environments. The aim of this study was to evaluate the psychometric properties of two novel, functionally oriented tests of lower limb position sense: the Gradient Discrimination Test (GradDT™) and Step-height Discrimination Test (StepDT™). Specific

objectives were to evaluate intra- and inter-rater reliability, discriminant validity and convergent validity and to determine sensitivity/specificity.

Methods

This is a reliability and validity study. A key catalyst to developing these tests was responding to the expressed views of people with stroke. Findings of a previous study [12], discussion amongst the patient, carer and public involvement group (PCPI) [27], as well as opinions of stakeholders working in stroke rehabilitation, contributed toward the development of the tests and study protocol. In response, the tests are designed to assess perceived difficulties with sensation in a functional context and are performed in full weight-bearing positions. Ethical approval was obtained from the UK NHS Health Research Authority NRES - Committee South Central – Berkshire B (15/SC/0191).

Participants

Participants were recruited from a convenience sample identified through local NHS community services and support groups. Eligibility criteria were: aged 18 and above, a stroke diagnosis confirmed via CT scan and clinical presentation, >3months post-stroke, able to independently stand and walk at least 10m indoors (with or without walking aid) and able to understand the information sheet or explanation of the research and provide informed consent. Potential participants were excluded if they had other neurological disease or co-morbidities/injuries that would affect mobility and/or foot sensory function. People with stroke were compared to healthy, age and gender matched controls.

Sample size was based on the work of Shoukri et al. [28]. For a 95% CI of 0.25 and a planned ICC of 0.8 (α=0.05), 32 participants were required. For inter-rater reliability, a study sample of 20 with two raters and a planned ICC of 0.8 (α=0.05) provides sufficient power for establishing a 95% CI of ~0.4 [29]. A sample size of 32 was sufficient for the test of

convergent validity to detect a correlation coefficient of 0.3 (power=0.85, α=0.05) and for discriminant validity to detect an effect size of 0.86 (power=0.85, α=0.05).

The Gradient Discrimination Test (GradDT™)

This novel test was developed to assess foot and ankle position sense during full weight bearing. The apparatus was designed to be as simple as possible to minimise expense of production, and maximise ease of administration. It comprised a bi-directional, rotating standing platform under the tested foot to produce a slope in the sagittal plane (Fig. 1). The axis of rotation was aligned with the lateral malleolus which is broadly speaking, the biomechanical axis of rotation when the foot is dorsiflexed and plantarflexed [30].

(insert Figure 1 around here)

The non-tested foot was positioned on a static, height adjustable horizontal platform, mirroring the height of the sloping (tested) platform when positioned at 0° relative to horizontal. The test foot was positioned on the rotating platform. Participants were required to transfer weight onto the non-tested foot, lift the tested foot clear of the platform and replace it once the platform had been adjusted to the new position by the assessor. This took up to three seconds. Platform angle was adjusted manually, by the assessor, using a laser cut acrylic “staircase” template, which allowed the gradient of the platform to be adjusted quickly, easily and quietly. Slope gradient and accuracy was supported through the use of a digital inclinometer attached to the platform (DigiPas, DWL 180S; resolution 0.05°). The base stimulus platform was 0° relative to horizontal and the comparator surface slope varied in increments of 0.5° up to a maximum of 10° dorsiflexion or plantarflexion. The participant was instructed to discriminate between the base platform and comparator platform, indicating which felt the most sloping. Upper limb support was available to aid with balance and participants were instructed to look straight ahead and to avoid looking at their feet.

The Step height Discrimination Test (StepDT™)

This novel test was designed to assess lower limb position sense discrimination thresholds. The apparatus was kept as simple as possible to minimise expense of production, and maximise ease of administration. The StepDT™ comprised an adjustable step, which contained a series of 6mm interlocking steps, allowing the height to be easily, quickly and quietly adjusted. The step height ranged from 100mm (base stimulus), which sits within the range of a standard kerb height, with the comparator step adjustable by 6mm increments up to a maximum step height of 154mm. The test involved the participant shifting weight onto the non-tested limb whilst the assessor passively placed the tested limb onto the step. The participant was instructed to discriminate between the height of base step and comparator step, indicating which felt the highest under the tested limb. The participant was provided with upper limb support to aid balance and instructed to look straight ahead through the test to avoid visual feedback of limb position.

A two-alternative forced choice design (2AFC) in combination with a 'one-up, three-down' staircase procedure [31] was employed in each test (fig 2).

Insert Fig 2 around here

The 2AFC staircase task is a psychophysical method where the aim is to determine at what point two (different) stimuli, cannot be accurately and consistently distinguished. The 2AFC aspect attempts to eliminate inconsistencies that can otherwise arise from different observers being more or less conservative when making subjective reports about ambiguous, near threshold stimuli. It is a fundamental methodology used in sensory science [32].

Applying the 2AFC design involved presenting two stimuli in quick succession (i.e. either gradient or step), in a series of increasingly difficult trials. Each trial included a base stimulus (A), and a changeable comparator stimulus (B). A and B were presented in random order (i.e. AB or BA) over the course of several trials. Stimuli were presented in a way that participants were unable to rely on any visual or auditory clues. Participants were required to indicate which stimulus most reflected the quality of interest. In the GradDT™, gradient slope was the quality of interest, with participants required to discriminate between base and comparator stimuli, indicating which has the greatest slope. In the StepDT™, the quality of interest was step height, with participants required to discriminate between base and comparator stimuli, indicating which step felt the highest using lower limb position sense. The staircase approach to the 2AFC design involved the systematic updating of the comparator stimulus depending on whether the participant was able to discriminate between the stimuli or not. The task became more difficult after three correct responses (i.e the difference between base and comparator stimuli was made smaller) or became easier following one incorrect response (i.e. difference between base and comparator was larger). This procedure is designed to converge over time on the threshold value that yields 79.4% correct performance. For example, if *p* is the probability of a positive response on a given trial, then *p* x *p* must equal 0.50. A three-down, one-up transformation leads to a performance target of 0.794 (i.e., *p* raised to power 3 = 0.50; the cube root of 0.50 is 0.794) [31]. The discrimination threshold was calculated from the average of four reversals (i.e. changes from a series of correct to a series of incorrect responses, or vice versa), triggered by the first incorrect response. A greater discrimination threshold in either test indicates worse somatosensory ability.

Both GradDT™ and StepDT™ are undertaken in standing with full weight bearing, which was importantly felt to reflect “real life” foot-ground sensorimotor interactions. During

testing, upper limb support for purposes of balance and safety was allowed via a hand rail/wall bars.

Procedures

Data collection was conducted in an outpatient hospital setting. Stroke participants (n=32) were tested with both tests on two occasions, between 7-14 days apart. The primary researcher (TG) was the rater on test session 1 and test session 2. A third testing session, involving 20 stroke participants, was completed by a physiotherapy assistant practitioner (PAP) 3-7 days after testing session 2. Training of the PAP included: test demonstration by the primary researcher, test practice under supervision, and the issuing of an operating procedures manual. Control participants (n=32) were tested on just one occasion. Concentration, working memory and attention were key requirements of the test so the testing environment was an enclosed, quiet room on each occasion. A small pilot study confirmed the GradDT™ and StepDT™ took 10-15 minutes each to complete and were well understood by people with stroke.

For the purposes of validity testing, in addition to the GradDT™ and StepDT™, further data was collected. This included: participant demographics and stroke characteristics, self-reported falls in the previous 3 months, sensory function using the Erasmus MC version of the Nottingham Sensory Assessment (EmNSA) [19], gait speed using the 10 metre timed walk test at fastest speed [33], and balance using the standing Forward Reach Test (FRT)[34]. Postural sway was quantified using centre of pressure velocity (COP velocity) measurement, reportedly most sensitive for detecting changes in balance abilities due to aging and/or neurological diseases [35] and was recorded using a Tekscan pressure mat (Matscan, Biosense medical, Essex, UK). There is currently no gold standard measure for establishing a dichotomous sensory status (impaired or not impaired) [14], and so self-report sensory status was used. This was determined through a yes/no response to whether, when asked,

participants felt they experience ‘difficulty judging or knowing the position of their feet without looking at them’, in line with the broad definition of proprioception [2].

Statistical analysis

Statistical analyses were performed using SPSS version 22.0. Data were summarised using frequencies and percentages, mean and standard deviation (SD) or median and inter-quartile range (IQR) as appropriate. Data distribution was assessed for normality using Shapiro-Wilks tests and assumed normally distributed when $p>0.05$. Data presented for the GradDT™ and StepDT™ represent gradient or height discrimination thresholds respectively, expressed in the original measurement unit (degrees or centimetres). The Just Noticeable Difference (JND) between base and comparator stimuli, in the StepDT™ is expressed as a percentage (%).

Necessary assumptions in reliability testing were accounted for [36]. These included: stability between testing sessions of participant sensory function and consistency in the testing situation (environment, test instructions, procedure). Both inter- and intra-rater reliability and agreement were analysed using intra class correlation coefficient ($ICC_{2,1}$) and Bland–Altman plots in line with recommendations [37]. Standard error of measurement (SEM) provided an indication of the score likely due to measurement error whilst Coefficient of repeatability (CoR), provides a score change, in the original measurement scale, which includes random and measurement error so represents a change reflective of a true/real change [38]. CoR is the value below which the absolute differences between two measurements would lie with 0.95 probability [36].

GradDT™ and StepDT™ performance of the paretic stroke foot and matched healthy control foot allowed for an evaluation of discriminant validity. A Mann Whitney U test was used to determine statistical significance between the groups ($p<0.05$). Convergent validity was evaluated by comparing the novel tests with an existing sensory measure. Although

there is no 'gold-standard' measure of somatosensation, the EmNSA is considered robust and clinically usable in neurological populations [14]. To evaluate convergent validity, the magnitude of the relationship between the proprioception subtest of the EmNSA and the GradDT™ and StepDT™ was determined using a Spearman's rank order correlation. The magnitude of the relationship between stroke participants' GradDT™ and StepDT™ performance and measures of gait speed, FRT, falls and COP velocity were evaluated using Spearman and Pearson correlational analysis where appropriate. Strength of correlations were interpreted using the classification where ≤0.29 = weak, 0.30- 0.49 = moderate and, ≥0.50 = strong [39].

The sensitivity and specificity of the GradDT™ and StepDT™, and for comparative purposes, the EmNSA, were evaluated using Receiver Operating Characteristic (ROC) curve analysis against the dichotomous variable of stroke participant self-report sensory impaired/not impaired. Sensitivity and specificity of a measure is used to quantify diagnostic ability, with sensitivity indicating the proportion of true positives that are correctly identified (i.e those who report sensory impairment), and specificity, the proportion of true negatives correctly identified (those without sensory impairment) by the tests [34]. The optimal cut-off score for each test in the original measurement scale is also useful in that it provides a score that can discriminate between those reporting sensory impairment and those not. This was calculated as least distance from the chance line or the Youden index (*J)* [40] using the formula: $J = (\sqrt{(1\text{-Sensitivity})^2 + (1\text{-Specificity})^2})$. This provides a cut-off point which optimizes test differentiating ability when equal weight is given to sensitivity and specificity [40].

Results

Thirty-two people with chronic stroke and 32 healthy age matched controls participated in the study (Table 1).

(insert table 1 around here)

Reliability and agreement

Intra- and inter-rater reliability data with mean (SD) discrimination thresholds for each of the tests, expressed in the original units of measurement, are presented in tables 2 and 3. Both variants of the GradDT™ and the StepDT™ demonstrated good to excellent intra- and inter-rater reliability.

(insert table 2 and 3 around here)

Bland-Altman plots demonstrated intra- and inter rater agreement for both variants of the GradDT™ and the StepDT™. Plots indicate the majority of scores fell within the limits of agreement (LOA: +/- 1.96SD) (Fig 3 a-c). In the StepDT™, (Fig 3-c) the mean of the differences (*d)* between rater 1 and rater 2 was +0.7cm with rater 2, on average, scored participants' step height discrimination threshold 0.7cm lower than rater 1 (i.e. participants consistently performed better when tested by rater 2) (Fig. 3-c). The line of equality/zero fell outside the 95% CI of mean difference (*d*, 95% CI 0.41cm - 0.91cm) (Fig 3c), suggesting a degree of systematic bias in participant performance when assessed by rater 2. Despite this, all plots were within LOA (+/- 1.96 SD) for the StepDT™.

(insert figure 3 around here)

Discriminant validity

Stroke participants had statistically significant higher discrimination thresholds (indicating worse position sense) than control participants in both tests (Table 4).

(insert table 4 around here)

Convergent validity

Data were not normally distributed, therefore Spearman's rank order correlation (rho) analysis was carried out to evaluate convergent validity. The GradDT™ in both dorsiflexion and plantarflexion conditions showed significant, moderate correlations ($r=-0.41$;$p<0.05$; $r=-0.47$; $p<0.01$ respectively) with tactile scores of the EmNSA, but weak and non-significant correlations with the proprioceptive scores of the EmNSA ($r=-0.17$;$p>0.05$; $r=-0.28$; $p>0.05$). The StepDT™ did not significantly correlate with the tactile or proprioceptive components of the EmNSA ($r=-0.14$, $r=-0.05$, $p>0.05$).

Table 5 shows Spearman's correlation coefficients between our novel measures, the EmNSA and functional outcome measures. The GradDT™ and StepDT™ had moderate to strong and significant correlations with gait speed indicating that those who had higher discrimination thresholds (i.e. poorer position sense) had slower walking speed (table 5).

(insert table 5 around here)

Moderate and significant correlations between $COP_{velocity}$ and the GradDT™ (DF and PF) conditions were also identified indicating that those with greater postural sway had higher ankle/foot position sense thresholds. The StepDT™ and both variants of the GradDT™, showed significant and strong correlations with the FRT. Only the StepDT™ significantly correlated with reported falls. The proprioception subtest of the EmNSA was not significantly correlated with any functional outcome measure (table 5).

Sensitivity & Specificity

Table 6 shows the results of the ROC curve analysis which evaluated the sensitivity and specificity of the GradDT™ and StepDT™ to identify self-reported sensory impairment/no impairment in stroke participants. The Area Under the Curve (AUC) statistic indicates excellent overall predictive ability for both tests.

(insert table 6 and fig 4 around here)

Both variants of the GradDT™ demonstrated a sensitivity of 79% and a specificity of 87%, with cut off points of 1.9° and 2.1° respectively. The 95% CI indicates the StepDT™ AUC value could be as low as 0.51 suggesting poor predictive value, or marginally better than chance (0.5). Using least distance and Youden index statistic, the optimal cut off point to predict subjective sensory impairment was deemed to be a step height discrimination threshold of 1.3cm, with high sensitivity and specificity (87% and 75% respectively) at this level. The proprioception subtest of the EmNSA had an AUC statistic of 0.54 (95% CI, 0.29-0.79) indicating poor predictive value. Figure 4 shows individual ROC curves for each test. The straight reference line running diagonally indicates a 0.5 probability of being diagnosed sensory impaired/not impaired i.e. no greater than chance. Curves to the left of the reference line indicate better diagnostic value than chance alone, whereas curves to the right of the line indicate worse diagnostic value. The closer the curve follows the top left corner, the better the diagnostic value.

Discussion and Conclusion

Here, two novel tests of lower limb position sense, the GradDT™ and StepDT™ were developed in response to a lack of functionally oriented, clinically feasible and sensitive measures of lower limb position sense. Their focus was derived through qualitative research [12] and their development supported by patient, carer, public and stakeholder involvement. This study evaluated the psychometric properties of these tests. The sensory-perceptual ability of the foot and ankle to discriminate surface gradient or slope was assessed during full weight-bearing using the GradDT™. The ability of the lower limb to discriminate the height of a step, without vision, was assessed with the StepDT™. The study results support the reliability and validity of these tests, and demonstrates their superior validity and sensitivity/specificity when compared to the proprioceptive component of the EmNSA.

The GradDT™ and StepDT™ have distinct advantages over several existing measures of lower limb position sense in that they target key functional areas related to stance, walking and stepping. They use a robust psychophysical testing approach to establish somatosensory discrimination thresholds, thereby assessing higher level processing of somatosensation [41], so they are potentially most relevant in (central) neurologically impaired populations. They employ an interval level of measurement meaning they allow for the degree of difference between scores, and show, in this sample, no floor or ceiling effects, thereby illustrating their potential responsiveness. SEM and CoR data provide the researcher/clinician with scores of random and measurement error, in the original measurement scale, providing real, true change; information which is critical for the monitoring of recovery and the evaluation of interventions. The GradDT™ and StepDT™ also demonstrated significant associations with functional measures of gait speed, static and dynamic balance and falls, so they may be of use in examining the relationship between functional ability, motor recovery and lower limb somatosensation.

However, there are limitations to this study and our tests. The testing of discriminative ability places demands on cognitive functions such as attention and working memory; functions which may also be impaired post stroke [42] and may be further confounded by factors such as fatigue – a known sequela of stroke [43]. Formal assessments of cognitive function and fatigue were not undertaken in this study, so the extent to which they influenced test outcome cannot be determined. Proprioception literally translates as "sense of self", so what constitutes proprioception is complex, ambiguous and frequently debated in the literature [44]. Several distinct sub-senses, in addition to sense of movement and position, have been proposed as proprioceptive sensations, such as sense of force, effort, and heaviness [2]. Our current novel measures do not therefore represent the full complement of proprioception, measuring just one aspect: sense of position. That these tests were evaluated in those capable of standing also limits their generalisability to the wider stroke population; further evaluation of our tests is required in other phases of stroke and across settings. In addition, the use of self-reported sensory status (impaired or not impaired) is potentially confounded by self-assessment ability, which may affect the sensitivity/specificity of these tests.

Inter-rater reliability of an outcome measure is crucial, particularly in long-term neurological populations who typically have multiple interactions with different health-care professionals over time. Reliability and agreement of both our novel tests was excellent and evaluated in accordance with recommended guidelines [37]. Issues around inter-rater reliability is commonly reported in measures of sensory testing [14] so these data are encouraging. For example, the reliability of knee position sense tests has been shown to vary substantially [45] whilst tests of ankle joint position sense (JPS) using motorised equipment [46, 47] report reliability values comparable to those in our study. Intra-rater reliability data from the GradDT™ and StepDT™ also compare favourably with the proprioception subtests

of traditional, clinical measures [14,17-19]. Bland Altman analysis revealed excellent intra- and inter rater agreement for the GradDT™ although systematic bias was evident in the StepDT™ with a consistently lower discrimination threshold recorded by the second rater. In addition, whilst the StepDT™ showed superior sensitivity/specificity to the EmNSA, the broad confidence interval (CI) questions its ability to predict self-reported sensory impairment. Both these data suggest that the involvement of assessor handling and the large amount of passive limb movement at hip and knee, may have confounded test performance.

There are several difficulties commonly associated with the measurement of proprioception. Firstly, a substantial proportion of proprioceptive information is processed by neural pathways which are not consciously mediated [2] and therefore inaccessible to subjective reporting and beyond measurement. Moreover, conscious and subconscious proprioceptive information is processed at every level of the CNS and by several structures making it difficult to speculate on how lesion location may affect perceptual ability of proprioception. Secondly, the links between proprioception and motor output mean the compartmentalisation or separation of proprioceptive function and motor function is challenging. Thirdly, sense of position and sense of movement have been shown to only weakly correlate [48] suggesting they are measuring different constructs. This may also explain why the GradDT™ and StepDT™ did not show significant correlations with the proprioceptive subtest of the EmNSA. The EmNSA assessed movement detection and movement direction with the participant in supine/sitting, whereas our novel measures assessed position sense in full weight-bearing.

There is currently no single gold-standard measurement tool capable of evaluating the diverse range of proprioceptive senses. Clinicians and researchers must therefore choose the measurement tool having first identified which aspect of proprioception is of interest, and in which joint/body part. The development of new tools which are able to differentiate the

proprioceptive senses will help to inform how individual sub-senses impact on movement and respond to treatment interventions.

The development and use of clinically feasibly and accurate proprioceptive measures, more closely aligned with functional ability, has been suggested by several reviewers [14,15,49,50]. Whilst current measures of position sense are geared toward identifying the presence of impairment, clinicians and patients are most concerned with addressing factors which impede function. This initial exploration of the psychometric qualities of our two novel tests suggests they hold great potential for use within both clinical and research settings.

References

1. Wolpert D. Pearson K & Ghez C. The Organization and planning of movement. In: Kandel E, Schwartz J, Jessell T, Siegelbaum S, Hudspeth A, editors. *Principles of neural science.* New York: McGraw Hill; 2013. p. 743-767
2. Proske U & Gandevia S. The proprioceptive senses: their roles in signalling body shape, body position and movement, and muscle force. *Physiol Rev* 2012; 92: 1651–1697
3. Latash M, Levin M Scholz J & Schoner G. Motor control theories and their applications. *Medicina (Kaunas*). 2010;46(6):382-92.
4. Bernardi N, Darainy M, & Ostry D. Somatosensory Contribution to the Initial Stages of Human Motor Learning. *J Neurosci*., 2015;35(42):14316 –14326
5. Ostry D, Darainy M, Mattar A, Wong J and Gribble P. Somatosensory Plasticity and Motor Learning. *J Neurosci*, 2010; 30(15):5384 –5393
6. Deshpande N, Metter EJ, Ferrucci L. et al. Validity of clinically derived Cumulative Somatosensory Impairment Index. Arch Phys Med Rehabil 2010;91:226-32.
7. Bowen C, Ashburn A, Cole et al. A survey exploring self-reported indoor and outdoor footwear habits, foot problems and fall status in people with stroke and Parkinson's. Journal of Foot and Ankle Research 2016; 9:39
8. Citaker S, Gunduz AG, Guclu M et al. Relationship between foot sensation and standing balance in patients with multiple sclerosis Gait Posture 2011;34:275-278
9. Gorst T, Rogers A, Morrison SC, Cramp M, Paton J, Freeman J & Marsden J. The prevalence, distribution, and functional importance of lower limb somatosensory impairments in chronic stroke survivors: a cross sectional observational study. *Disab Rehabil* 2018, DOI: 10.1080/09638288.2018.1468932
10. Connell, L. Lincoln, N., and Radford, K. Somatosensory impairment after stroke: frequency of different deficits and their recovery. Clin Rehabil 2008; 22(8): 758–67
11. Patel A, Duncan P, Lai S, Studenski S. The relation between impairments and functional outcomes poststroke. Arch Phys Med Rehabil. 2000; 81(10):1357-63p

12. Gorst T, Lyddon A, Marsden J et al Foot and ankle impairments affect mobility and balance in stroke (FAiMiS): the views of people with stroke. *Disabil Rehabil*. 2016;38(6):589-96
13. Tyson S, Crow L, Connell L, Winward C, & Hilier S. Sensory Impairments of the lower limb after stroke; A pooled analysis of individual patient data. *Topics in Stroke Rehabilitation* 2013; 20(5): 441-449
14. Connell, L & Tyson, S. Measures of sensation in neurological conditions; a systematic review. *Clin Rehabil* 2012; 26: 68
15. Hillier S, Immink M, Thewlis D. Assessing proprioception: a systematic review of possibilities. *Neurorehabil Neural Repair* 2015;29:933–49
16. Yalcin E, Akyuz M, Onder B, et al Position Sense of the hemiparetic and non-hemiparetic ankle after stroke: is the non-hemiparetic ankle also affected? *Eur Neurol* 2012; 68:294-299
17. Winward CE, Halligan PW and Wade DT. The Rivermead Assessment of Somatosensory Performance (RASP): standardisation and reliability data. *Clin Rehabil* 2002; 16: 523–33
18. Lincoln NB, Jackson JM and Adams SA. Reliability and revision of the Nottingham Sensory Assessment for Stroke Patients. *Physiotherapy* 1998; 84: 358–65.
19. Stolk-Hornsveld F, Crow JL, Hendriks EP, van der Baan R and Harmeling-van der Wel BC. The Erasmus MC modifications to the (revised) Nottingham Sensory Assessment: a reliable somatosensory assessment measure for patients with intracranial disorders. *Clin Rehabil* 2006; 20: 160–72
20. Fugl-Meyer A, Jaasko L, Leyman I, Olsson S and Steglind S. The post-stroke hemiplegic patient: a method for evaluation of physical performance. *Scand J Rehab* Med 1975; 7: 13–31.
21. Lynch EA, Hillier SL, Stiller K, Campanella RR, Fisher PH. Sensory retraining of the lower limb after acute stroke: a randomized controlled pilot trial. *Arch Phys Med Rehabil* 2007;88:1101-7.
22. Lin J, Hsueh I, Sheu C and Hsieh C. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. *Clin Rehabil* 2004; 18: 391–97.
23. Elangovan N, Herrmann A, Konczak J. Assessing proprioceptive function: evaluating joint position matching methods against psychophysical thresholds. *Phys Ther*. 2014;94:553–561.
24. Wingert J, Burton H, Sinclair R, Brunstrom J and Damiano D. Joint-position sense and kinesthesia in cerebral palsy. *Arch Phys Med Rehabil* 2009; 90: 447–53
25. Lin S.-I. Motor function and joint position sense in relation to gait performance in chronic stroke patients. *Arch Phys Med Rehabil* 2005; 86(2): 197–203.
26. Leibowitz N, Levy N, Weingarten S, Grinberg Y, Karniel A, Sacher Y, et al. Automated measurement of proprioception following stroke. *Disab Rehabil* 2008; 30: 1829–36.
27. INVOLVE (2012) Briefing notes for researchers: involving the public in NHS, public health and social care research. INVOLVE, Eastleigh NIHR
28. Shoukri M, Asayli M, Donner A. Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research* 2004; 13: 1:21

29. Doros G and Lew R. Design Based on Intra-Class Correlation Coefficients. *American Journal of Biostatistics* 2010;1 (1): 1-8,
30. Palastanga N & Soames R. *Anatomy and Human Movement: Structure and Function*. 6th edition 2012, Churchill Livingstone: Elsevier
31. Leek M. Adaptive procedures in psychophysical research. *Perception & Psychophysics* 2001, 63 (8), 1279-1292
32. Bi J. *Sensory Discrimination Tests and Measurements: Statistical Principles, Procedures and Tables*. 2006. Blackwell Publishing
33. Bohannon RW, Andrews AW, Thomas MW. Walking speed: reference values and correlates for older adults. *J Orthop Sports Phys Ther*. 1996;24(2):86-90
34. Weiner, D. K., Duncan, P. W., et al. Functional reach: a marker of physical frailty. *J Am Geriatr Soc* 1992; 40(3): 203-207
35. Masani A, Vette A, Abe M. Center of pressure velocity reflects body acceleration rather than body velocity during quiet standing *Gait & Posture* 39; 2014: 946–952
36. Bland M. *An introduction to medical statistics*. 4th edition. 2015. Oxford; Oxford University Press.
37. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *Journal of Clinical Epidemiology* 2011; 64: 96-106
38. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P (2013) The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS ONE* 8(9): e73990
39. Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic
40. Youden, W.J. Index for rating diagnostic tests. *Cancer*. 1950; 3: 32–35
41. Borstad A & Nichols-Larsen D. Assessing and Treating Higher Level Somatosensory Impairments Post Stroke. *Topics in Stroke Rehabilitation* 2014; 21(4): 290-295
42. Crichton S, Bray B, McKevitt C, Rudd A & Wolfe C. Patient outcomes up to 15 years after stroke: survival, disability, quality of life, cognition and mental health. *Neurol Neurosurg Psychiatry* 2016; 87:1091–1098
43. Acciarresi M, Bogusslavsky J & Paciaroni M. Post-Stroke Fatigue: Epidemiology, Clinical Characteristics and Treatment. *Eur Neurol* 2014;72:255–261;
44. Krewer C, Van de Winckel A, Elangovan N et al. Commentary on: "Assessing proprioception: A critical review of methods" by Han et al. *Journal of Sport and Health Science* 2016; 5 :91–92
45. Smith T, Davies L, Hing C. A systematic review to determine the reliability of knee joint position sense assessment measures. *The Knee* 2013; 20: 162–169
46. Sun, W Qipeng Song, Bing Yu, Cui Zhang & Dewei Mao. Test–retest reliability of a new device for assessing ankle joint threshold to detect passive movement in healthy adults. *Journal of Sports Sciences*, 2015; 33(16): 1667-1674
47. Ko S-K, Simonsick E, Deshpande N & Ferrucci. Sex-specific age associations of ankle proprioception test performance in older adults: results from the Baltimore Longitudinal Study of Aging. *Age and Ageing* 2015; 44: 485–490
48. de Jong A, Kilbreath SL, Refshauge KM, Adams R. Performance in different proprioceptive tests does not correlate in ankles with recurrent sprain. *Arch Phys Med Rehabil* 2005;86:2101–5.

49. Carey L, Lamp G, & Turville M. The State-of-the-Science on somatosensory function and its impact on daily life in adults and older adults and following stroke: A scoping review. *OTJR: Occupation, Participation and Health*. 2016; 36(2s):27s-41s
50. Suetterlin K & Sayer A. Proprioception: where are we now? A commentary on clinical assessment, changes across the life course, functional implications and future interventions. *Age and Ageing* 2014; 43: 313–318

Tables

Table 1. Stroke and control participant demographic and clinical characteristics

Characteristics	Stroke (n=32)	Control (n=32)
Age, years, mean (SD)	70 (9)	70 (7)
Gender n (%)		
Male	22 (69)	19 (59)
Female	10 (31)	13 (41)
Modified Rankin Score n (%)		
1	12 (38)	-
2	10 (31)	-
3	10 (31)	-
Stroke type n (%)		
Ischaemic	25 (78)	-
Haemorrhagic	7 (22)	-
Time since stroke		
Months, mean (SD)	22 (18)	-
Side most affected n (%)		
Right	18 (56)	-
Left	14 (44)	-
Number of falls reported n (%)		
0	16 (50)	27 (84)
1	6 (19)	3 (9)
2	3 (9)	2 (7)
3	5 (16)	0
>4	2 (9)	0
Gait speed, mean m/s (SD)	1.1 (0.6)	1.7 (0.4)
FRT cm, mean (SD)	23.7 (9.1)	34.0 (6.6)
$COP_{velocity}$ mm/s mean (SD)	9.6 (18.2)	1.0 (12.4)

m/s, metres per second; SD, Standard Deviation; FRT, Functional Reach Test; cm, centimetres; COP, centre of pressure; mm/s, millimetres per second

Table 2. Intra-rater reliability scores for the GradDT™ and StepDT™ (stroke participants, n=32)

	Intra-rater Reliability (n=32)					
Measure	Test 1 (T1)	Test 2 (T2)	Mean (T1 &T2)	SEM	$ICC_{(2,1)}$ (95% CI)	CoR
GradDT™ (PF) threshold degrees (°) mean (SD)	3.1 (1.9)	3.4 (2.1)	3.2 (2.0)	0.60	0.91 (0.82-0.96)*	1.6
GradDT™ (DF) threshold degrees(°) mean (SD)	2.9 (1.9)	3.1 (1.9)	3.0 (1.9)	0.63	0.89 (0.79-0.95)*	1.7
StepDT™ threshold cm mean (SD)	2.5 (1.2)	2.4 (1.2)	2.4 (1.2)	0.30	0.95 (0.90-0.97)*	0.75

GradDT, Gradient Discrimination Test; StepDT, Step-height Discrimination Test; PF, plantarflexion DF, dorsiflexion cm, centimetres; SD, Standard Deviation; SEM, Standard error of measurement; $ICC_{(2,1)}$ Intraclass Correlation Coefficient model 2,1; CI, Confidence Interval; CoR, Coefficient of Repeatability
*P<0.001

Table 3. Inter-rater reliability scores for the GradDT™ and StepDT™ (stroke participants, n=20)

	Inter-rater Reliability (n=20)					
Measure	Rater 1 (R1)	Rater 2 (R2)	Mean (R1 &R2)	SEM	$ICC_{(2,1)}$ (95% CI)	CoR
GradDT™ (PF) threshold degrees (°) mean (SD)	2.8 (1.7)	2.5 (1.8)	2.6 (1.7)	0.45	0.93 (0.82-0.97)*	1.2
GradDT™ (DF) threshold degrees(°) mean (SD)	2.5 (1.5)	2.6 (1.9)	2.5 (1.7)	0.48	0.92 (0.79-0.97)*	1.3
StepDT™ threshold cm mean (SD)	2.0 (1.0)	1.4 (1.1)	1.7 (1.0)	0.38	0.85 (0.64-0.94)*	1.1

GradDT, Gradient Discrimination Test; StepDT, Step-height Discrimination Test; PF, plantarflexion DF, dorsiflexion; cm, centimetres; SD, Standard Deviation; SEM, Standard error of measurement; $ICC_{(2,1)}$ Intraclass Correlation Coefficient model 2,1; CI, Confidence Interval; CoR, Coefficient of Repeatability
*P<0.001

Table 4. Stroke and control participant performance in the GradDT™ and StepDT™.

Measure	Stroke (n=32)	Control (n=32)	p	Odds Ratio (95% CI)
GradDT™ (PF) threshold degrees (°) median (IQR, range)	3.1° (2.8, 8.3)	1.5° (1.1, 2.5)	<0.001	3.57 (1.63-7.69)
GradDT™ (DF) threshold degrees (°) median (IQR, Range)	3.0° (2.4, 8.3)	1.2° (1.0, 2.5)	<0.001	4.76 (2.08-9.09)
StepDT™ threshold cm median (IQR, Range)	1.8cm (2.4, 3.6)	1.2cm (0.6, 1.8)	<0.001	6.67 (2.38-20.00)

GradDT, Gradient Discrimination Test; StepDT, Step-height Discrimination Test; PF, plantarflexion DF, dorsiflexion; IQ, Inter-quartile range; cm, centimetres; CI Confidence Interval.

Table 5. Spearman's correlations between novel measures, EmNSA and functional outcome measures (stroke participants (n=32)

Measure	Gait speed	FRT	COPv	Fall Incidence
GradDT™ (PF)	-0.40*	0.57**	-0.44*	0.17
GradDT™ (DF)	-0.47**	0.57*	-0.43*	0.11
StepDT™	-0.60**	0.59**	0.19	0.56**
EmNSA (prop)	-0.04	0.18	-0.17	-0.20

FRT, Functional Reach Test; COPvelocity, Centre of Pressure velocity; EmNSA, Erasmus modified Nottingham Sensory Assessment; *p<0.05, **p<0.01

Table 6. Sensitivity and specificity analysis of GradDT™, StepDT™ and EmNSA

Test	AUC	SE	95% CI	*p* value	Cut Off Point	Sensitivity (%)	Specificity (%)
GradDT™							
Dorsiflexion	0.83	0.07	0.68-0.98	0.005	1.9°	79	87
Plantarflexion	0.87	0.06	0.74-0.99	0.002	2.1°	79	87
StepDT™	0.75	0.12	0.51-0.98	0.03	1.3cm	87	75
EmNSA Proprioception	0.54	0.125	0.29-0.79	0.71	6.50	46	50

AUC, Area Under the Curve; SE, Standard Error; CI, Confidence Interval; GradDT, Gradient Discrimination Test; StepDT, Step-height Discrimination test; EmNSA, Erasmus modified Nottingham Sensory Assessment

Figures

Fig. 1. Experimental set up of the Gradient Discrimination Test (GradDT™)

Fig 2. Procedural algorithm of two alternative forced choice design (2AFC) using a three down, one –up staircase procedure

Fig. 3 a-c Bland Altman plots of threshold scores showing difference between rater 1 and rater 2 scores plotted against mean threshold scores for rater 1 and rater 2 across: a) GradDT™ (PF); b) GradDT™ (DF); and c) StepDT™. Horizontal (small dashed) lines indicate upper and lower levels of agreement (LOA; +/- 1.96 Standard Deviation (SD); large dashed lines indicate 95% Confidence Interval of mean difference (*d*)

Fig 4. Receiver Operating Characteristic (ROC) curve illustrating sensitivity/specificity of novel measures and proprioceptive component of Erasmus modified Nottingham Sensory Assessment against dichotomous self-report of impaired sensation/not impaired sensation