Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder

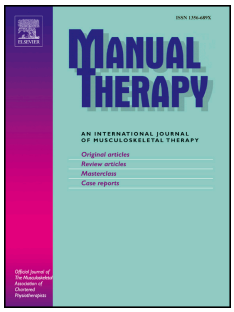
Tyros, I.; Soundy, Andrew; Heneghan, Nicola

DOI:
10.1016/j.math.2016.06.006

Document Version
Peer reviewed version

Citation for published version (Harvard):
Tyros, I, Soundy, A & Heneghan, N 2016, 'Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder: intra- and inter-rater reliability study', Manual Therapy, vol. 25, pp. 81-86. https://doi.org/10.1016/j.math.2016.06.006

Accepted Manuscript



Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder: intra- and inter-rater reliability study

Mr I. Tyros, Senior Physiotherapist, Dr A. Soundy, Lecturer Physiotherapy, Dr N.R. Heneghan, Lecturer Physiotherapy

PII: S1356-689X(16)30637-3

DOI: 10.1016/j.math.2016.06.006

Reference: YMATH 1871

To appear in: *Manual Therapy*

Received Date: 21 September 2015

Revised Date: 6 June 2016

Accepted Date: 7 June 2016

Please cite this article as: Tyros I, Soundy A, Heneghan N, Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder: intra- and inter-rater reliability study, *Manual Therapy* (2016), doi: 10.1016/j.math.2016.06.006.

Title

Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder: intra- and inter-rater reliability study

Authors

Tyros I, Soundy A, Heneghan NR,

Affiliations

Mr Isaak Tyros
Senior Physiotherapist
Physio 1st LTD
Birmingham
B68 0NP

Dr Andrew Soundy
Lecturer Physiotherapy
School of Sport, Exercise and Rehabilitation Sciences
College of Life and Environmental Sciences
University of Birmingham
Edgbaston
Birmingham
B15 2TT

Dr Nicola R Heneghan
Lecturer Physiotherapy
School of Sport, Exercise and Rehabilitation Sciences
College of Life and Environmental Sciences
University of Birmingham
Edgbaston
Birmingham
B15 2TT

Corresponding author and reprint requests:

Mr Isaak Tyros

Tel: +44 (0) 7821 559311

isaak.tyros@gmail.com

Keywords:

Intra-rater, Inter-rater reliability; tuning fork; Whiplash Associated Disorders; vibration testing

Vibration sensibility of the median nerve in a population with chronic whiplash associated disorder: intra- and inter-rater reliability study

Whiplash Associated Disorders (WAD) grade II are the most prevalent group of whiplash patients seen on a regular basis by musculoskeletal physiotherapists. Impairment of vibration sensibility may be an early indicator of nerve pathology and it has previously been demonstrated in individuals with chronic WAD symptoms utilising vibrameters. A less expensive option, such the tuning fork (TF) may assist with these measures, but research regarding its measurement properties is lacking.

Objectives: To investigate the intra- and inter-rater reliability of vibration sensibility of the median nerve in chronic WAD II (CWADII).

Methods: A double blinded, within day intra- and inter-rater reliability study was undertaken. A convenience sample of 26 individuals (8 males, 18 females, age mean 29.9 ±10.0 years) with CWADII was recruited. Exclusion criteria: WAD I, III & indications of neuropathic pain. Vibration attenuation times were recorded from skin innervated by the median nerve (thenar eminence).

Results: Descriptive statistics (mean scores) and reliability statistics [intraclass correlation coefficient ($ICC_{2,1}$) and Bland and Altman limits of agreement] were undertaken with p=0.05. Almost perfect intra-rater reliability (ICC: 0.972-0.955) and inter-rater reliability (ICC: 0.983) were identified. Confidence Intervals (CI) for inter-rater reliability were 95% CI: -1.461 to -0.056.

Conclusions: Almost perfect reliability scores across intra- and inter-rater reliability were found. This provides evidence that, with a standardised testing protocol the TF can be a highly reliable means of vibration sensibility testing. Future studies assessing the validity of the TF in different WAD populations may provide further information about the usefulness of this protocol.

INTRODUCTION

Whiplash Associated Disorders (WAD) are a significant socioeconomic problem which is increasing throughout the industrialised world (Rushton et al., 2011). Recent economic data (2010) puts the annual economic cost of WAD, relating to management, lost market and household productivity, at $128 billion in the USA alone (Blincoe et al., 2014).

Clinically, WAD can be classified as grades (0-IV), depending on the severity of presentation, with more severe presentations being higher grades (Spitzer et al., 1995); a system widely adopted in practice (Nederhand et al., 2000; Sterling, 2004). WAD II represent the most common group of patients (93.4%) experiencing neck pain along with stiffness or tenderness, and musculoskeletal signs (Sterling, 2004; Rushton et al., 2011). This group is differentiated from WAD III where patients also present with neurological signs such as decreased or absent reflexes, muscle weakness and sensory deficits (Spitzer et al., 1995).

Individuals with chronic WAD (grade II) may also demonstrate signs of local and/or generalized mechanical and cold hyperalgesia as well as altered sympathetic nervous system activity (Kasch et al. 2001; Sterling et al., 2003; Sterling 2004).

The studies of Chien et al., (2008b & 2009) were some of the first studies that incorporated Quantitative Sensory Testing (QST) in the assessment of chronic WAD populations and reported coexistence of generalised sensory hypersensitivity and hypoesthetic changes. Such hypoesthetic changes (loss of sensitivity to vibration, electrical and thermal stimuli) indicate a dysfunction of Aβ, Aδ and C nerve fibres and are suggestive of a neuropathic component in the presentation of chronic WAD (Chien et al., 2008b; Chien et al., 2009; Chien & Sterling 2010). Similar findings have been found in other musculoskeletal conditions (patellofemoral pain and chronic diffuse upper limb pain) and may be also associated with central inhibitory processes related to the duration of the nociceptive input (Jensen et al. 2008; Tucker at al. 2007; Chien and Sterling 2010).

Altered vibration sensibility has also been described as an early indicator of nerve pathology in several conditions (Halonen et al., 1986; Greening et al. 2003) specifically indicating dysfunction of Aβ afferent fibres and the Pacinian corpuscles (Dyck, et al., 1987; Reis & Moro, 2012). According to Leak (1998), vibration sensibility testing is rarely used clinically, due to the suggested subjectivity, lack of investigated reliability (e.g. tuning fork) or inconvenience of the available testing instruments (e.g. vibrameter).

Vibration sensibility is assessed through QST, although it is just one of a number of a group of tests representing measures of all relevant features of the somatosensory system (Rolke et al. 2006). The QST enables clinicians to identify the underline mechanism of pain disorders (Edwards et al., 2005). Its validity has been demonstrated by Rolke et al., (2006) who suggested that QST is useful for building the somatosensory profile of patients with suspected neuropathic pain.

Quantitative Sensory Testing can assist the clinician establish the severity of a clinical presentation and lead to more accurate management choices, however the method employs costly equipment, is time consuming and needs further evaluation before it can be efficiently applied in clinical practice (Chien & Sterling 2010). All current studies that have utilised QST within WAD populations have used a vibrameter (Somedic AB, Sweden) to detect vibration disappearance thresholds (VDT). For clarity, VDT is operationally defined as the time until the perception of vibration disappears. Whilst the vibrameter remains the gold standard for the assessment of VDT (James & Scott 2012), it is costly and not available in most clinical settings. The tuning fork (TF) offers an inexpensive, convenient and widely available alternative to measure VDT in clinical practice, although little evidence exists for its reliability.

Reliability of the Tuning Fork (128Hz)

Only two studies were identified that have examined the reliability of the TF (128Hz) on asymptomatic populations, and provide conflicting information as to the reliability of the tool. Botez et al., (2009) demonstrated that VDT can be measured with almost perfect (Landis & Koch, 1977) intra- and inter-rater reliability (ICC 0.79-0.92 and ICC 0.82-0.95 respectively), while O'Conaire et al., (2011) demonstrated moderate (Landis & Koch, 1977) inter-rater reliability (ICC 0.52). The two studies differed in their designs with Botez et al., (2009) using single blinding and a non-standardised method for production of vibration stimulus and O'Conaire et al., (2011) employing double blinding and a standardised method for producing vibration amplitudes with a novel device adjunctive to the TF. Both studies used an asymptomatic population and failed to control the application pressure of the TF. To date there are no studies that have examined the reliability of the TF (128Hz) in a symptomatic population.

Clinically, utilisation of tools that are able to provide reliable responses between raters, and within raters are necessary. Considering all the above, there is a need for more inter-and intra-rater reliability studies related to the TF (128Hz).

Aims

To investigate the intra- and inter-rater reliability of the TF measuring vibration sensation on a median nerve distribution in participants with chronic WAD grade II (CWAD II).

MATERIALS AND METHODS

Design and sample

A double blinded, cross-sectional, within day intra- and inter-rater reliability study. A convenience sample of twenty six (n = 26) participants (university staff and students) with CWAD II symptoms were recruited from the University of Birmingham, based on a power calculation (Walter et al., 1998) designed for studies incorporating three consecutive measurements, with the Intraclass Correlation Coefficiency (ICC) range between 0.6-0.8 and the significance level, α=.05. According to Walter et al. (1998) a protocol with three measurements (n=3) would be convenient as further measurements could

potentially lead to a fatigue effect and an undesirable aversion effect. Using these above specific indices, and with the minimum ICC value set at 0.6 ($\rho 0$) and the maximum set at 0.8 ($\rho 1$), it was estimated that a total of k=26 participants would be required. Based on this power calculation, 26 participants were recruited.

Inclusion criteria: age 18-69 years, a history of a neck injury, and presentation of CWAD II symptoms for at least 6 months. Exclusion criteria: presentation of neck complaint of pain, stiffness, or tenderness only (WAD I), idiopathic neck pain, history of severe neck trauma (fracture, dislocation, WAD IV), indications of neurological deficits (WAD III), and indications of neuropathic pain (S-LANSS score of >12), currently receiving active clinical management. Two raters and an assistant were randomly recruited from a convenient population of four clinical musculoskeletal physiotherapists on a postgraduate programme. The assistant's role was to record the measurements and keep raters and patients blinded from the results. All measurements took place in a laboratory setting to control for external factors and allow similar environmental conditions (Hicks, 2004).

Ethics

The study was approved by the School of Health Sciences of the University of Birmingham (Ethics Reference number: PGT_1314_037).

Equipment

A 128 Hz Tuning Fork (Ragg Gardiner Brown Co, 11 Furnace Hill, Sheffield, S3 7AF, England) was used to measure VDT. Such devices are widely used in clinical practice and research (Fillyaw et al., 1989; Richardson, 2002; O'Neill et al., 2006; Botez et al., 2009; O'Conaire et al., 2011). A pen cap of 14mm diameter was used to establish standardised vibration amplitudes. A "Fastime Zero 1" stopwatch was used for the recording each subject's VDT.

In order to better characterise the sample, two questionnaires were utilised. The S-LANSS questionnaire was used to identify individuals with lower symptom scores who reflect the common patients attending clinical practice and to possibly avoid those with high scores (>12) that may have neuropathic pain. With specificity ranges from 75% to 80% and sensitivity from 74% to 78% the S-LANSS is considered a valid and reliable self-report outcome measure (Bennett et al. 2005). Neck Disability Index (NDI) was used to evaluate evidence of disability. The NDI has demonstrated a high degree of reliability and internal consistency (Vernon & Mior, 1991; McCarthy et al., 2007; Walton et al., 2009) and is used to enable classification of WAD (Sterling, 2004).

Testing Protocol

Similarly to previous studies and current protocols for trials, VDT was calculated from the mean of three consecutive measurements (Leak, 1998; Rolke et al., 2006; O'Conaire et al., 2011; Tesarz et al. 2013). This study's testing protocol aimed to standardise the application pressure of the TF by using its own weight as the only pressure applied on the tested area (Heneghan & Rushton, 2012 – Verbal

Communication). Training was necessary to ensure raters would develop the precision needed for holding the TF upright without adding any further pressure on the tested area.

Vibration attenuation times were recorded from skin innervated by the median nerve over the thenar eminence; selected due to the frequency of median nerve's symptoms in WAD where C5, C6 and C7, segments are the most commonly affected (Sizer et al., 2004; Chien et al., 2009; Sterling et al., 2010).

Procedure

Prior to the measurements raters undertook two sessions of training to familiarise themselves with the procedure and the technical handling of the TF. The testing room environment was quiet and clear of distractions with a standardised temperature of 22 - 24^{o}C (Hilz et al., 1998; Rolke et al., 2006). Participants had to assume a supine lying position on an examination bed with their cervical spine in neutral position and arms held in a relaxed position, with slight shoulder abduction, elbows fully extended and wrist in neutral (Figure 1). Whilst in position, participants wore a blindfold (Figure 1) for achieving maximal concentration during the measurements (Hilz et al., 1998; Botez et al., 2009). A rater from the standardised sitting position (Figure 1) placed the TF, perpendicular to the skin of the thenar eminence of the dominant hand. The VDT was recorded based on a participants' verbal response that they no longer felt the sensation of vibration. All VDT scores were recorded by the assistant to ensure blinding of both raters and participants from the results. The assistant would start the stopwatch simultaneously with raters' counting down to ''three'' and stopped it when the participant pronounced ''now''. The test was always first performed by Rater 1. There was one minute rest between the three measurements, with five minutes wash out period, before Rater 2 repeated the same procedure. Raters were both blinded to their own previous findings and the findings of the other rater. Also, both raters were blinded to any other clinical information that was not intended to be provided as part of the testing procedure.

Figure 1 – Position of rater and participant

Statistical Analysis

Descriptive statistics presented a preliminary impression of the data including the variability of measurements between the two raters. The p value was set at 0.05. The ICC (model 2,1) was selected for the calculation of intra- and inter-rater reliability values. Model 2 was selected as the measurements on each participant were carried out by the same set of randomly selected raters (Muller and Buttner, 1994; Sim and Wright, 2000). Normality of data was assessed using Kolmogorov-Smirnov test (Sim and Wright, 2000). A Bland Altman limits of agreement plot (Bland and Altman, 1986) was used to visually quantify the measurement error. To assist with interpretation of the reliability analysis and precision, the 95% confidence intervals, 95% limits of agreement and the standard error were calculated (Atkinson and Nevill, 1998; Weir, 2005). Data were analysed using SPSS version 21.

RESULTS

Twenty six participants (8 males, 18 females) were recruited, mean age 29.9 ±10.0 years. NDI (mean of 16.18 ±0.11) and S-LANSS (mean of 3.53 ±3.20) questionnaire scores are reflective of a sample with mild neck disability and no indications of neuropathic pain.

Normality of data

Rater 1 demonstrated a significant deviation from normality (p=0.04) for the first rating, but not for rating 2 (p>0.05) or 3 (p>0.05). Rater 2 demonstrated no significant deviation across all 3 ratings (p>0.05). Standardised z-scores for kurtosis (-1.24) and skewness (1.16) for the first rating of Rater 1 fell within an acceptable range (between ±1.96; Field, 2009).

Descriptive statistics

Vibration disappearance thresholds (in seconds) mean ± SD for measurements of Rater 1: (39.18 ± 9.24, 38.37 ± 8.82 and 37.90 ± 9.63) and Rater 2: (37.88 ± 9.32, 37.73 ± 8.78 and 37.55 ± 8.34).

Intra-rater reliability: For Rater 1 the intra-rater ICC value was 0.972 whereas for Rater 2 was 0.955. Data were statistically significant for both raters as (p<0.001, F=37.1 for Rater 1 and F=21.5 for Rater 2).

Inter-rater reliability: Analysis of the data using the $ICC_{2,1}$ average measures, demonstrated inter-rater agreement of 0.983 (95% CI: 0.971 – 0.991). Data were statistically significant as (p<0.00001, F=59.5).

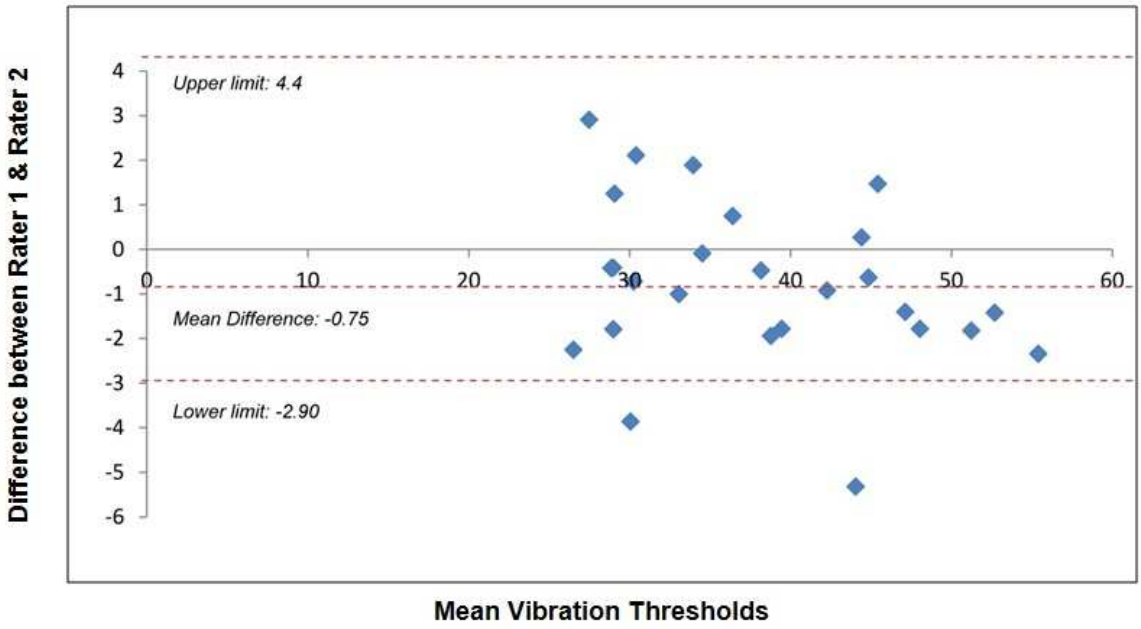
Precision analysis: Values for SEM and 95% Confidence Intervals are shown in Table 1. The Bland Altman plot (Figure 2) depicts the 95% limits of agreement and these are between 4.4 and -2.90 with a mean of -0.75.

Table 1 – *Precision Analysis for inter-rater reliability*

95% Limits of Agreement		SEM (true SEM[b])	95% Confidence Intervals	
Upper Limit	Lower Limit	0.358 (0.702)	Lower Interval	Upper Interval
4.415	-2.899		-1.461	-0.056

[b] True SEM = 1.96*SEM

Figure 2 – Bland Altman 95% Limits of Agreement - Inter-rater comparison



DISCUSSION

This is the first study to investigate vibration sensibility reliability using a clinically available tool within a sample of individuals with WAD. The design and protocol used in the current study addressed important limitations, which are thought to affect the reliability of the TF (Gregg, 1951; Bleeker, 1986; Leak, 1998). These limitations traditionally include the following: inconsistent amplitude production; amplitude loss throughout transfer of the TF to the patient, and the amount of pressure application of the TF on the tested area (Gregg, 1951; Bleeker, 1986; Leak, 1998; Botez et al., 2009; O'Conaire et al., 2011). The current protocol utilised some of the characteristics used by O'Conaire's et al., (2011). It is believed that addressing these limitations explains the high intra- and inter-rater reliability found in the current study.

Application pressure of the tuning fork

Both Botez et al., (2009) and O'Conaire et al., (2011) recognised that the most significant limitation of their studies was failure to control the application pressure of the TF and highlighted the importance of such a standardisation for improving reliability scores of this method. This is in contrast with Leak's (1998) conclusion that vibration amplitude was not significantly affected by the application pressure of the TF. However, it should be taken into account that in Leak's (1998) study, vibration testing was performed on a solid surface which has different mechanical properties from those of human soft tissue (Nielsen, 1975; Bleeker, 1986). According to Goldberg and Lindblom (1979), the spreading and continuation of vibration in a tested tissue can be significantly affected by the mechanical impedance of the tissue. Therefore, Leak's (1998) findings can not be correlated with studies looking into vibration testing on human soft tissues and failure to control the application pressure of the TF should be considered as an important limitation.

This study demonstrated that following a short training, standardisation of TF's application pressure can be achieved by clinicians with or without experience with the tool. Rater 1 had previous experience utilising this study's protocol (having participated in a pilot study). Rater 2 used this testing protocol for the first time but reliability scores were high amongst them.

To date, no other study has previously described a protocol that standardises the application pressure of the TF which was the most commonly reported limitation in the current literature (Leak, 1998; Richardson, 2002; Botez et al., 2009; O'Conaire et al., 2011).

Standardisation of the amplitude

Non-standardised production of vibration amplitudes has been highlighted as an issue for use of a TF. A number of different techniques have been described for the initiation of vibration amplitudes, among which are: a maximal strike of TF's prongs on a hard surface (Leak, 1998; Richardson, 2002); and a maximal compression of TF's prongs followed by a sudden release (Aaserud et al., 1990). In the study of Botez et al., (2009) raters "clanged" the prongs of the TF against their thenar eminences in order to initiate vibration, but again this technique is non-standardised across participants. These techniques are considered quite subjective due to the inability of force clibration used by raters for the production of the amplitudes (Leak, 1998). Transfer of the TF onto the testing area during testing would also result in some loss of vibration amplitude (Leak, 1998). Unfortunately, O'Conaire's protocol cannot be generalised, as his novel device that controlled for these limitations is not currently available.

This study's protocol was based on O'Conaire's et al., (2011) research but with a simpler approach involving a pen cap to standardise the amplitude. The prongs of the TF were inserted into the empty cap and by the time this cap was removed it produced a standardised vibration. The aim of this study was to provide clinicians with a much simpler and equally standardised protocol, by incorporating a cluster of economical and obtainable materials. Moreover, pens cap are inexpensive, and widely available.

Positioning

Neither the study of Botez et al., (2009) nor the study of O'Conaire et al., (2011) provided details regarding the position of raters while taking the measurements. The position of raters could have affected the application pressure of the TF and therefore the reliability of the test. In order to avoid this limitation, raters in this study adopted a sitting posture in order to standardise the angles of their upper limbs while using the TF and avoid any additional effort produced by other postures like standing or bending over.

In both the studies of Botez et al., (2009) and O'Conaire et al., (2011), participants were examined from a sitting position adopting current practices of bedside examination. However, previous studies have demonstrated that a sitting position can be related to respiratory changes, increased lumbar lordosis and activation of postural muscles (Andersson et al., 1979; Lin et al., 2006). Given the psychophysical nature of vibration testing (Hilz et al., 1998) it could be considered that a lying position is more suitable for allowing participants to achieve maximal relaxation and maintaining the focused concentration,

which is essential in QST (Rolke et al., 2006). The utilisation of a lying position in this study may further explain the high reliability results.

Timing vibration thresholds

For timing, VDT Botez et al. (2009) used a wristwatch to stimulate a typical clinical setting, but this practice has not been justified by previous studies and adds to the subjectivity of this method (Leak, 1998). O'Conaire's et al., (2011) innovative device controlled this limitation, as the device's stopwatch was initiated simultaneously with the release of the prongs from its grip. However, in the absence of O'Conaire's et al., (2011) novel device, the inclusion of an assistant in this study's protocol was imperative for handling the stopwatch, while raters were performing the measurements. Taking into account the existence of electronic applications in tablets and mobile devices that can handle complex functions, the need for an assistant can likely be waived.

Limitations

One of the limitations of this study is the study sample consisting of students and University staff with WAD not receiving treatment for their complaints. This sample cannot be extrapolated to clinical practice as these patients are not seeking care for their complaints. Also, the sample of this study (WAD grade II), with low ratings on the S-LANSS, may not always reflect samples in clinical practice, thus requiring further verification. An experimental paradigm was performed to ensure maintained concentration of participants and assist with the psychophysical measurements employed in this study, which is essential for QST testing (Rolke et al., 2006). However, this sample may respond differently to individuals with WAD who usually have concentration and attention problems which can affect the results of testing. The use of a convenience sample raises concerns regarding generalisability of the results to other populations (Hicks, 2004). A further limitation regarding the generalisability of the results would be the homogeneity of the sample. Scores of NDI (mean = 16.18% ±0.11) and S-LANSS questionnaire (mean = 3.53 ±3.20), reflected a sample with mild neck disability and likely without indication of pain of predominantly neuropathic origin. The results can only be generalised to chronic WAD II populations with mild symptoms and not to those with moderate or severe level of disability or different WAD classifications.

Implications for clinical practice

This study took into consideration previous calls for the development of economical and reliable approaches to vibration testing (Pestronk et al., 2004). The TF has been shown to be a valuable measurement tool for use in a controlled laboratory environment given it is portable, inexpensive, quick and easy to use. Until now, controversy around TF's reliability and subjectivity were the main reasons for its rare use in clinical practice. The results of this study suggest that the TF when used with this new standardised protocol (Table 2) can be highly reliable and therefore could become not only a convenient but also an economical alternative to electronic testing (vibrameter) for the assessment of VDT. Clinicians after a short training can appropriately use the TF with the standardised protocol and assess VDTs quickly and reliably.

Use of the TF with this protocol could re-introduce vibration sensibility testing in clinical practice and/or be incorporated within QST. However, this protocol needs to be tested in a clinical setting and investigated for validity before being advocated for use in clinical practice.

Table 2 - Key points of the testing protocol

Rater's sitting position to minimise effort
TF to be hold (not pushed) perpendicularly to skin for its weight to standardise the pressure application
Removal of pen cap in situ
To be tested on soft tissue instead of bony surfaces
Close eyes to enhance concentration on the vibratory stimulus
Timing: An electronic device (stopwatch with a countdown) or application can synchronise this better

Future Research

Future research should focus on evaluation of the validity of the TF with this protocol in different clinical settings, such as primary care (and secondary care settings). This protocol's validity can also be compared with the vibrameter which remains the gold standard for measuring vibration thresholds (James & Scott, 2012) in both symptomatic and asymptomatic populations. Future studies should also assess the validity and the reliability of this method on other WAD grades (e.g. WAD grade III), in order to further inform assessment methods for this condition.

REFERENCES

Aaserud, O., Juntunen, J. and Matikainen, E. (1990) Vibration sensitivity thresholds: methodological considerations. **Acta Neurologica Scandinavica** 82: 277-283

Andersson, G.B., Murphy, R.W. and Ortengren R. et al. (1979) The influence of backrest inclination and lumbar support on lumbar lordosis. Spine, 4 (1): 52-58

Atkinson, G. and Nevill, A.M. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. **Sports Medicine,** 26 (4): 217-238

Bennett, M.I., Smith, B.H. and Torrance, N. et al. (2005) The S-LANSS Score for Identifying Pain of Predominantly Neuropathic Origin: Validation for Use in Clinical and Postal Research. The Journal of Pain, 6 (3): 149-158

Bland, J.M. and Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. **Lancet,** 1 (8476): 307-310

Bleeker, M.L. (1986) Vibration perception thresholds in entrapment and toxic neuropathies. **Journal of Occupational Medicine,** 28 (10): 991-994

Blincoe L.J., Miller, T.R. and Zaloshnja E. et al. (2014). *The economic and societal impact of motor vehicle crashes, 2010.* Washington, DC, National Highway Traffic Safety Administration

Botez, S.A., Liu, G. and Logigian, E. et al. (2009) Is the bedside timed vibration test reliable? **Muscle & Nerve,** 39 (2): 221-223

Chien, A., Eliav, E. and Sterling, M. (2008b) Whiplash (grade II) and cervical radiculopathy share a similar sensory presentation: an investigation using quantitative sensory testing. **Clinical Journal of Pain,** 24 (7): 595-603

Chien, A., Eliav, E. and Sterling, M. (2009) Hypoaesthesia occurs with sensory hypersensitivity in chronic whiplash? Further evidence of a neuropathic condition. **Manual Therapy,** 14 (2): 138-146

Chien, A. and Sterling, M. (2010) Sensory hypoaesthesia is a future of chronic whiplash but not idiopathic neck pain. **Manual Therapy,** 15 (1): 48-53

Dyck, P.J., Bushek, W. and Spring, E.M. et al. (1987) Vibratory and cooling detection thresholds compared with other tests in diagnosing and staging diabetic neuropathy. **Diabetes Care,** 10 (4): 432-440

Edwards, R.R., Sarlani, E. and Wesselmann, U. et al. (2005) Quantitative assessment of experimental pain perception: multiple domains of clinical relevance. **Pain,** 114 (3): 315-319

Field, A. (2009) **Discovering Statistics Using SPSS.** 3rd ed. London: Sage.

Fillyaw, M.J., Badger, G.J. and Bradley, W.G. et al. (1989) Quantitative measures of neurological function in chronic neuromuscular diseases and ataxia. **Journal of the Neurological Sciences,** 92 (1): 17-36

Goldberg J., Lindblom U. (1979) Standardised method of determining vibratory perception thresholds for diagnosis and screening in neurological investigation. **Journal of Neurology, Neruosurgery and Psychiatry** 42 (9): 793-803

Greening, J., Lynn, B. and Leary R. et al. (2003) Sensory and autonomic function in the hands of patients with non-specific arm pain (NSAP) and asymptomatic office workers. **Pain,** 104 (1-2): 275-281

Gregg, E. C. (1951). Absolute measurement of the vibratory threshold. **Archives of Neurology and Psychiatry (Chicago),** 66: 403-411

Halonen, P., Halonen, J.P. and Landen, P.A. et al. (1986) Vibration perception thresholds in shipyard workers exposed to solvents. **Acta Neurologica Scandinavica,** 73 (6): 561-565

Hicks, C. (2004) **Research Methods for Clinical Therapists: Applied project design and analysis.** 4th ed. Edinburgh: Churchill Livingstone

Hilz, M.J., Axelrod, F.B. and Hermann, K. et al. (1998) Normative values of vibratory perception in 530 children, juveniles and adults aged 3-79 years. **Journal of Neurological Sciences,** 159 (2): 219-225

James, G. and Scott, C. (2012) Vibration testing: A pilot study investigating the intra-tester reliability of the Vibrameter for the Median and Ulnar nerves. **Manual Therapy,** 17 (4): 369-372

Jensen, T., Kvale, A. and Baerheim, A. (2008) Is pain in patellofemoral pain syndrome neuropathic? **Clinical Journal of Pain,** 24 (5): 384-394

Kasch, H., Stengaard-Pedersen, K. and Arendt-Nielsen, L, et al. (2001) Pain thresholds and tenderness in neck and head following acute whiplash injury: a prospective study. **Cephalalgia,** 21 (3): 189-197

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. **Biometrics,** 33 (1): 159-174

Leak, S.V. (1998) Measurement of physiotherapists' ability to reliably generate vibration amplitude and pressures using a tuning fork. **Manual Therapy,** 3 (2): 90-94

Lin, F., Parthasarathy, S., and Taylor, S.J. et al. (2006) Effect of different sitting postures on lung capacity, expiratory flow, and lumbar lordosis. **Archives of Physical Medicine and Rehabilitation,** 87 (4): 504-509

McCarthy, M.J.H., Grevitt, M.P. and Silcocks, P. et al. (2007) The reliability of the Vernon and Mior neck disability index, and its validity compared with the short form-36 health survey questionnaire. **European Spine Journal,** 16 (12): 2111-2117

Muller, R. and Buttner, P. (1994) A critical discussion of intraclass correlation coefficients. **Statistics in Medicine,** 13 (23-24): 2465-2476

Nederhand, M.J., IJzerman, M.J. and Hermens, H.J. et al. (2000) Cervical Muscle Dysfunction in the Chronic Whiplash Associated Disorder Grade II (WAD-II). **Spine,** 25 (15): 1938-1943

Nielsen, V. K. (1975). The vibration stimulus. Effects of viscous-elastic resistance of skin on the amplitude of vibrations. **Electroencephalography and Clinical Neurophysiology**, 38: 647-652

O' Conaire, E., Rushton, A. and Wright, C. (2011) The assessment of vibration sense in the musculoskeletal examination: Moving towards a valid and reliable quantitative approach to vibration testing in clinical practice. **Manual Therapy,** 16 (3): 296-300

O'Neill, J., McCann, S.M. and Lagan, K.M. (2006) Tuning fork (128 Hz) versus Neurothesiometer: a comparison of methods of assessing vibration sensation in patients with diabetes mellitus. **International Journal of Clinical Practice,** 60 (2): 174-178

Pestronk, A., Florence, J. and Levine, T. et al. (2004) Sensory exam with a quantitative tuning fork. **Neurology,** 62 (3):461-464

Reis, P. and Moro, A. (2012) Preventing Rsi/Wruld: use of esthesiometry to assess hand tactile sensitivity of slaughterhouse workers. **Work,** 41 (1): 2556-2562

Richardson, J.K. (2002) The Clinical Identification of Peripheral Neuropathy Among Older Persons. **Archives of Physical Medicine and Rehabilitation,** 83 (11): 1553-1558

Rolke, R., Magerl, W. and Campbell, K.A. et al. (2006) Quantitative sensory testing: a comprehensive protocol for clinical trials. **European Journal of Pain,** 10 (1): 77-88

Rushton A, Wright C, and Heneghan N, et al. (2011) Physiotherapy rehabilitation for whiplash associated disorder II: a systematic review and meta-analysis of randomised controlled trials. **British Medical Journal Open,** 1:e000265. doi:10.1136/bmjopen-2011-000265

Sim, J. and Wright, C. (2000) **Research in Health Care: Concepts, Designs and Methods.** London: Stanley Thornes

Sizer, P.S., Pourbaugh, K. and Phelps, V. (2004) Whiplash Associated Disorders: Pathomechanics, Diagnosis, and Management. **Pain Practice,** 4 (3): 249-266

Spitzer, W.O., Salmi, L.R., and Cassidy, J.D. et al. (1995) Scientific monograph of the Quebec Task Force on Whiplash- Associated Disorders: redefining 'whiplash' and its management. **Spine,** 15 (20): 1S–73S

Sterling, M., Jull, G. and Vicenzino, B. et al. (2003) Sensory hypersensitivity occurs soon after whiplash injury and is associated with poor recovery. **Pain,** 104 (3): 509-517

Sterling, M. (2004) A proposed new classification system for whiplash associated disorders: implications for assessment and management. **Manual Therapy,** 9 (2): 60-70

Sterling, M., Ashley, P. and Chan, C. et al. (2010) Cervical lateral glide increases nociceptive flexion reflex threshold but not pressure or thermal pain thresholds in chronic whiplash associated disorders: A pilot randomized controlled trial. **Manual Therapy,** 15 (2): 149-153

Teasell, R.W., McClure, J.A. and Walton, D. et al. (2010) A research synthesis of therapeutic interventions for whiplash-associated disorder: Part 1 – overview and summary. **Pain, Research and Management,** 15 (5): 287-294

Tesarz, J., Gerhardt, A. and Schommer, K. et al. (2013) Alterations in endogenous pain modulation in endurance athletes: An experimental study using quantitative sensory testing and the cold-pressor task. **Pain,** 154 (7): 1022-1029

Tucker, A., White, P. and Kosek, E. at al. (2007) Comparison of vibration perception thresholds in individuals with diffuse upper limb pain and carpal tunnel syndrome. **Pain,** 127 (3): 263-269

Vernon, H. and Mior, S. (1991) The neck disability index: a study of reliability and validity. **Journal of Manipulative and Physiological Therapeutics,** 14 (7): 409–415

Walter, S.D., Eliasziw, M. and Donner, A. (1998) Sample size and optimal designs for reliability studies. **Statistics in Medicine,** 17 (1): 101-110

Walton, D.M., Avery, S. and Blanchard, A. et al. (2009) Measurement properties of the Neck Disability Index: A systematic review. **Journal of Orthopaedic & Sports Physical Therapy,** 39 (5): 400-417

Weir, J.P. (2005) Quantifying test-retest reliability using the Intraclass Correlation Coefficient and the SEM. **Journal of strength and conditioning Association,** 19 (1): 231-240

ACKNOWLEDGEMENTS

Authors would like to thank the Chartered Society of Physiotherapy (CSP), Elsevier and the Musculoskeletal Association of Chartered Physiotherapists (MACP), for supporting this study with educational bursaries. Authors would also like to thank staff and students of the University of Birmingham for their participation and support.

Highlights

- Whiplash associated disorder II (WAD II) are seen on a regular basis by musculoskeletal physiotherapists
- The tuning fork is an inexpensive clinically available tool to assess vibration sensibility
- 'Almost perfect' intra-rater reliability (ICC: 0.972-0.955) and inter-rater reliability (ICC: 0.983) were identified
- The tuning fork provides an inexpensive reliable assessment tool for vibration sensibility in CWAD II.