Manual Therapy 16 (2011) 563–572

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Original article

Clinical relevance vs. statistical significance: Using neck outcomes in patients with temporomandibular disorders as an example

Susan Armijo-Olivo [a,b,c,*], Sharon Warren [c], Jorge Fuentes [c,d,e], David J. Magee [c]

[a] Alberta Research Centre for Health Evidence (ARCHE), Faculty of Medicine and Dentistry, Aberhart Center, Room No. 8414, 11402 University Avenue, Edmonton, Alberta, Canada T6G 2J3
[b] Research Center, Faculty of Rehabilitation Medicine, University of Alberta, 3-50 Corbett Hall, Edmonton, Alberta, Canada T6G 2G4
[c] Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, 3-48 Corbett Hall, Edmonton, Alberta, Canada T6G 2G4
[d] Faculty of Graduate Studies and Research, University of Alberta, Edmonton, Alberta, Canada T6G 2G4
[e] Department of Physical Therapy, Catholic University of Maule, Talca, Chile

ARTICLE INFO

Article history:
Received 18 November 2010
Received in revised form
16 April 2011
Accepted 9 May 2011

Keywords:
Clinical relevance
Neck outcomes
Temporomandibular disorders
Effect size
Minimal important difference

ABSTRACT

Statistical significance has been used extensively to evaluate the results of research studies. Nevertheless, it offers only limited information to clinicians. The assessment of clinical relevance can facilitate the interpretation of the research results into clinical practice. The objective of this study was to explore different methods to evaluate the clinical relevance of the results using a cross-sectional study as an example comparing different neck outcomes between subjects with temporomandibular disorders and healthy controls. Subjects were compared for head and cervical posture, maximal cervical muscle strength, endurance of the cervical flexor and extensor muscles, and electromyographic activity of the cervical flexor muscles during the CranioCervical Flexion Test (CCFT). The evaluation of clinical relevance of the results was performed based on the effect size (ES), minimal important difference (MID), and clinical judgement. The results of this study show that it is possible to have statistical significance without having clinical relevance, to have both statistical significance and clinical relevance, to have clinical relevance without having statistical significance, or to have neither statistical significance nor clinical relevance. The evaluation of clinical relevance in clinical research is crucial to simplify the transfer of knowledge from research into practice. Clinical researchers should present the clinical relevance of their results.

1. Introduction

Most of the results of research in general and health research have used statistical significance in order to demonstrate effectiveness of an intervention, differences among groups in some variables of interest, or associations between variables. Statistical significance is based on hypothesis testing (Kirk, 1996). The null hypothesis states that there is no difference between groups or that an independent variable does not have an effect on the dependent variable. The alternative hypothesis states that groups are different or that an independent variable does have an effect on the dependent variable. After conducting the research, the statistical analysis provides one with the "*p*" value which indicates the strength of the evidence against the null hypothesis. Thus, statistical significance analysis only provides a dichotomous answer: it may or may not be statistically significant (in other words we have enough evidence against the null hypothesis or not) (Sterne and Smith, 2001). Therefore, statistical significance does not offer an indication of how important the result of the study is (Thompson, 1999; Ogles et al., 2001; Millis, 2003).

Statistical significance can also provide misleading results. A statistical difference between groups could be found if the sample size is large and if the intersubject variability is low, even though the difference between groups is small to be considered clinically important (Millis, 2003). Some authors have argued that tests of statistical significance are not generally useful and instead confidence intervals (CIs) and measures of effect size should be the main focus of research findings since they can provide more complete information regarding the magnitude of the association between variables, changes after a treatment, or differences between groups (Olejnik and Algina, 2000; Sterne and Smith, 2001). For example, CIs contain all of the information provided by a significance test in addition to a range of values within which the true difference is

* Corresponding author. Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, 3-50 Corbett Hall, Edmonton, Alberta, Canada T6G 2G4. Tel.: +1 780 492 1694/7247; fax: +1 780 492 1626.
E-mail addresses: sla4@ualberta.ca, susanarmijo@gmail.com (S. Armijo-Olivo), sharon.warren@ualberta.ca (S. Warren), jorgef@ualberta.ca (J. Fuentes), david.magee@ualberta.ca (D.J. Magee).

1356-689X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.math.2011.05.006

likely to lie. This information facilitates understanding of the "magnitude of the effect" by researchers and clinicians and offers a richer source of information in addition to the simple yes/no dichotomy of hypothesis testing (McNeely and Warren, 2006).

A result can be clinically relevant but might be neglected if statistical significance was not attained due to small sample sizes and high intersubject variability. Clinical relevance (also called clinical significance) assessment indicates whether the results are meaningful or not. In this way the evaluation of clinical relevance can provide more interesting results for health care clinicians as well as clients receiving care, facilitating the transfer of knowledge into clinical practice (Musselman, 2007). Some authors in the areas of education (Kirk, 1996; Carnine, 1997) as well as health research (Millis, 2003; Musselman, 2007) have urged that research findings be reported in language that is familiar to practitioners. With the advancement of health care and the introduction of evidence based practice, researchers need to provide information regarding their research that can be used in clinical practice and demonstrate an impact in health care and clinical decisions. The information of "*p*" values is insufficient to achieve these requirements and because it provides insufficient and limited information, clinical researchers needed to present the clinical relevance of their results to help busy clinicians with interpretation.

Some methods to determine clinical relevance have been created in order to provide clinicians, clients and policy makers with standards of meaningful change. The most common and used methods to determine clinical relevance are "distribution-based methods" and "anchor-based methods". Distribution-based methods are based on the statistical distribution and the psychometric properties of the outcomes. The calculation of the effect size, the minimal important difference (MID), and the standard error of measurement are examples of distribution-based methods to evaluate clinical relevance. Anchor-based methods involve the clients' perspective in the assessment of clinical relevance and are used prospectively.

Clinical relevance is generally evaluated as a result of an intervention; however, clinical relevance can also be assessed in other types of research such as cross-sectional studies. In these studies, patients and controls are assessed on certain variables of interest and it is important to know if the differences found between groups are in fact clinically meaningful. The interpretation of a score on a certain outcome in a cross-sectional study is performed by comparing the values obtained with those found in a reference population. Unfortunately, most of the outcomes used in clinical research lack "normative or reference values" to establish "normality" of health status. Thus the interpretation of clinical relevance for results in this type of research is uncertain and difficult to make for the general practitioner. Therefore, other methods for assessing clinical relevance in cross-sectional studies need to be used in the absence of normative values for the outcomes of interest. In addition, information regarding clinical relevance for neck outcomes is lacking and clinicians have difficulty to interpret results from research studies. Thus, the objectives of this paper were (1) to explore and analyze different methods to evaluate the clinical relevance of the results using a cross-sectional study as an example comparing different neck outcomes between subjects with temporomandibular disorders (TMD) and healthy controls and (2) to discuss different issues regarding clinical relevance and statistical significance when interpreting these results.

2. Methods

2.1. Assessing clinical relevance

2.1.1. Sample data

The data used for this example was obtained from a large study investigating the involvement of cervical spine in patients with TMD. Details regarding this study are described elsewhere (Armijo-Olivo, 2010; Armijo-Olivo et al., 2010b). The general description of the sample is as follows.

2.1.1.1. Subjects. Subjects with TMD who attended the Orofacial Pain/TMD clinic at University of Alberta and healthy students and staff at the University of Alberta were recruited for this study. The inclusion/exclusion criteria for healthy and subjects with TMD and characteristics of the subjects have been described extensively elsewhere (Armijo-Olivo, 2010; Armijo-Olivo et al., 2010b). Included subjects signed an informed consent in accordance with the University of Alberta's policies on research using human subjects. One hundred and fifty-four participants (154) provided data for this example. From these 154 subjects, 50 subjects were healthy, 56 subjects had myogenous temporomandibular disorders (TMD) and 48 subjects had mixed TMD.

2.1.2. Procedures

Demographic and clinical data were collected on all subjects who satisfied the inclusion criteria.

In addition, subjects with TMD (i.e. myogenous and mixed TMD) were compared with healthy subjects in the following variables:

1. Head and cervical posture,
2. Maximal cervical muscle strength,
3. Endurance of the cervical flexor muscles,
4. Endurance of the cervical extensor muscles during the neck extensor muscle endurance test (NEMET), and
5. EMG activity of the cervical flexor muscles during the craniocervical flexion test (CCFT).

More details about data collection and set up of the experiments can be found elsewhere (Armijo-Olivo, 2010; Armijo-Olivo et al., 2010b). A brief description of the analyzed variables is as follows.

2.1.2.1. Head and cervical posture. Head and neck posture were measured in a lateral photograph, taken with the head in the self-balanced position (Cooke and Wei, 1988; Sandoval et al., 1999). Four angles were measured on the photographs: (1) eye-tragus-horizontal, (2) tragus-C7-horizontal, (3) pogonion-tragus-C7, and (4) tragus-C7-shoulder using Alcimagen software® (Gadotti et al., 2005; Cesar et al., 2006). All of the measurements were performed by a single trained rater, a dentist specializing in orthodontics, blinded to the subjects' group status, following the same procedure for all photographs. More details about the procedure can be found in Armijo-Olivo et al. (Armijo-Olivo et al., in press-a).

2.1.2.2. Maximal cervical flexor strength. Maximal cervical flexion strength was measured with the subjects in supine position using a device attached to a plinth and connected to a visual feedback screen. This device contained a load cell to register the strength generated by the subject during the procedure. The average value of strength of the 2 contractions registered was used as the maximal voluntary contraction (MVC). More details can be found in Armijo-Olivo et al. (Armijo-Olivo et al., 2010c).

2.1.2.3. Endurance of the cervical flexor muscles. The endurance of the cervical flexor muscles was performed in the same position (i.e. *supine position*) and using the same equipment described for the evaluation of the MVC. After performing the MVC, each subject was asked to perform two submaximal cervical flexion contractions at 25% MVC, 50% MVC, and 75% MVC, keeping the chin retracted, and to maintain these contractions as long as possible using a visual display for feedback of the force output. The holding time during the cervical flexion movement at different levels of contraction was

registered and analyzed. The test was stopped when (1) the subject could not maintain the desired target strength level (i.e. percentage MVC) determined for the test, or (2) the subject complained (self-reported) of an unacceptable pain during the test or the training stage (Armijo-Olivo et al., 2010a).

2.1.2.4. Endurance of the cervical extensor muscles during the neck extensor muscle endurance test (NEMET). The endurance of the neck extensor muscles was measured using the neck extensor muscle endurance test (NEMET). Subjects were asked to maintain a prone position on a plinth with the head and neck unsupported over the end of the plinth with the arms alongside the trunk. Endurance holding time was measured with a stopwatch after removing the neck support and asking the subject to hold the position of the head steady with the chin retracted and the cervical spine horizontal to the floor (Armijo-Olivo, 2010). The test was discontinued if: (Lee et al., 2005)

1. The subject complained of fatigue or pain in the neck or if the subject complained of intolerable pain in another part of the body (i.e. thoracic spine, interscapular region, low back)
2. The subject could not maintain the head in the horizontal position. This was determined when the lights were "on" for longer than 5 s on more than 5 occasions.
3. The subject lost more than 5° of upper cervical retraction for more than 5 s as measured by the level goniometer located in the subjects' head (LIC rehab Vardrum, Solna, Sweden).

2.1.2.5. EMG activity of the cervical flexor muscles during the craniocervical flexion test (CCFT). The performance of the superficial cervical flexor muscles was evaluated through the craniocervical flexion test (CCFT) (Falla et al., 2003). The CCFT required each subject to perform the craniocervical flexion movement in five progressive stages of increasing pressure (between 22 and 30 mmHg) with the aid of a visual feedback device. The electromyographic activity of the sternocleidomastoid and anterior scalenes (right and left) was collected during the CCFT. To obtain a measure of EMG amplitude, maximum root mean square (RMS) was calculated for 4 s during the 10-s submaximal contractions for each muscle while doing the CCFT using IGOR Pro5.1‡ and was expressed a percentage of the 3 s EMG activity obtained during the MVC normalization procedure (Armijo-Olivo, 2010; Armijo-Olivo et al., in press-b).

2.2. Analysis

The statistical analysis of each one of these variables has been extensively presented elsewhere (Armijo-Olivo, 2010; Armijo-Olivo et al., 2010a,c; Armijo-Olivo et al., in press-a,b,c). Only some of the results are presented here to illustrate statistical significance vs. clinical relevance. The focus of this present study was to evaluate and highlight the clinical relevance of the results. In this study, evaluation of clinical relevance of the results of each of the variables when comparing subjects with TMD and healthy subjects was performed based on the distribution-based method using the effect size (ES) (Cohen) (Perera et al., 2006), minimal important difference (MID),(Guyatt et al., 2002; Lemieux et al., 2007) and clinical judgement (Musselman, 2007). A brief description of these concepts and methods to calculate the clinical relevance of the study results, used in this study, is as follows:

2.2.1. Effect size (ES)

Effect size has been defined by Cohen (Cohen, 1988) as: "the degree to which the phenomenon is present in the population", so the larger the effect size, "the greater the degree to which the phenomenon under study is manifested" (Cohen, 1988) p. 10. According to Ogles (Ogles et al., 2001), ES can provide information regarding the magnitude of association between variables as well as the size of the difference between groups. Commonly, the ES is calculated by dividing the difference between group mean scores (i.e. control group and patient group; control group and intervention group; pre intervention scores and post intervention scores) by the standard deviation at baseline, by the standard deviation of the control group, or by the pooled standard deviation of the 2 groups (Portney and Watkins, 2000). The following formula is generally used to calculate the effect size (Portney and Watkins, 2000)

$$\overline{\mathrm{ES}} = \frac{\overline{X_{G1}} - \overline{X_{G2}}}{S_{\mathrm{pooled}}}$$

where:

ES = effect size
X_{G1} = mean group 1
X_{G2} = mean group 2
S_{pooled}: pooled standard deviation (SD)

To calculate the pooled standard deviation, the following formula is used

$$S_{\mathrm{pooled}} = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}}$$

where:

S_1 = SD group 1
S_2 = SD group 2
n_1 = sample size for group 1
n_2 = sample size for group 2

The magnitude of the effect size has been interpreted as an index of clinical relevance (Kirk, 1996; Musselman, 2007). The larger this effect size index, the larger the difference between groups and the larger the clinical relevance of the results (Musselman, 2007). Since effect size is a measure that can be applicable to all kinds of research designs and statistical models (Cohen, 1988; Kirk, 1996), the use of effect size as a measure of clinical relevance can be used not only in research based on interventions but also in other types of non-experimental research (e.g. correlational, cross-sectional) (Callahan and Reio, 2006).

Cohen described 0.2, 0.5, and 0.8 as small, moderate and large effect sizes respectively (Cohen, 1988). However, these values are used only as guidance to make decisions (Callahan and Reio, 2006; Kirk, 2007). Thus, it was decided for this study that an arbitrary and more conservative cut off would be used. Therefore, an effect size of $\mathrm{ES} \geq 0.4$ was considered clinically relevant since this effect or difference could represent a moderate effect which might be of interest for clinical practice (Cohen, 1988).

2.2.2. Minimal important difference (MID)

The MID has been defined as "the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management" (Guyatt et al., 2002). The concept of MID is important for clinicians since it can help guide treatment decisions. Also, it helps researchers with study design and sample size calculation which is crucial for

ensuring trial quality. Thus, calculation of MID provides a starting point for interpreting clinical relevance of trials results.

One method of determining the MID is by multiplying the effect size of the difference obtained between groups considered as important (0.2 or 0.5 ES according to Cohen) by the pooled baseline standard deviation between the 2 groups ($S_{\text{pooled baseline}}$) (Lemieux et al., 2007). The following formula is used:

$$\text{MID} = 0.2 \times S_{\text{pooled baseline}}$$

or

$$\text{MID} = 0.5 \times S_{\text{pooled baseline}}$$

For interpretation purposes, a mean difference between groups that is higher than the MID can be considered as clinically relevant (Lemieux et al., 2007; Musselman, 2007).

2.2.3. Clinical judgement

Evaluation of clinical relevance can be performed using clinical experience. This is the simplest way to evaluate clinical relevance when the clinician has a vast experience in the field and knows the outcomes (Musselman, 2007). Although this way of evaluating clinical relevance is subjective and depends on the expertise of the evaluator, in some cases, clinical judgement overrides the calculations of clinical relevance. According to Sloan (Sloan, 2005), "clinical opinion should trump statistical theory". However, in many cases clinicians do not have the criterion to determine clinical relevance and therefore, calculation of clinical relevance can help to develop a criterion or a guideline.

2.2.4. Clinical relevance: final decision

In order to make a final decision regarding clinical relevance for the analyzed outcomes in this present study, all criteria (i.e. effect size, MID and clinical judgement) for determining clinical relevance were used. A result was considered "clinically relevant" if both the calculated ES was $\geq$0.40 and both of the calculated MIDs were lower than the mean difference obtained between groups (i.e. healthy and TMD groups). A result was considered "potentially clinically relevant" if ES was small–moderate or moderate and only one of the MIDs was lower than the mean difference obtained between groups. A result was considered "not clinically relevant" when neither the ES nor the MIDs were accomplished (i.e. ES $<$ 0.40 and MIDs were higher than mean difference obtained between groups) or because the clinical judgement determined it as not clinically relevant.

3. Results

Mean differences and 95% confidence intervals between groups in the variables of interest, as well as values for clinical relevance based on different methods (i.e. effect size, and MID) are described in Tables 1–3.

A summary of the specific results of each one of the variables is as follows.

3.1. Head and cervical posture

The only angle considered statistically significantly different between groups was the eye-tragus-horizontal angle. The mean difference between subjects with myogenous TMD and healthy subjects in this angle was 2.6°. The calculated effect size for this difference was 0.46 (approaching to a moderate effect size). The MID was 1.08° and 2.70° using 0.2 and 0.5 effect sizes respectively for the calculation. Although the calculated values of clinical relevance demonstrated an important finding in 2 of the comparisons (moderate effect sizes, see Table 1), the differences found between subjects with TMD and healthy individuals were considered not to be clinically relevant based on clinical judgement.

3.2. Maximal cervical flexor strength

Maximal cervical flexor muscle strength was not statistically significantly different between patients with TMD and healthy subjects. Average differences in maximal cervical flexor muscle

Table 1
Clinical relevance assessment head and cervical posture in patients with temporomandibular disorders when compared with healthy subjects.

Outcome	Mean difference (°)	Confidence interval for mean difference		Pooled SD	Effect size (ES) or standardized mean difference	ES based on SD of control group	Interpretation ES	MID (0.2) = 0.2 × pooled SD	MID (0.5) = 0.5 × pooled SD	Final decision clinical relevance
		Lower	Upper							
Head and cervical posture										
Eye-tragus-horizontal angle: mixed TMD vs. healthy (°)	1.5	−0.7	3.7	5.6	0.3	0.3	SES	1.1	2.8	NCR[a]
Eye-tragus-horizontal angle: myogenous TMD vs. healthy (°)	2.6	0.5	4.7	5.4	0.5	0.5	MES	1.1	2.7	NCR[a]
Tragus-C7-horizontal: mixed vs. healthy (°)	−1.7	−4.0	0.5	5.6	−0.3	−0.3	SES	1.1	2.8	NCR[a]
Tragus-C7-horizontal: myogenous vs. healthy (°)	0.6	−1.5	2.6	5.3	0.1	0.1	SES	1.1	2.7	NCR[a]
Pogonion-tragus-C7: mixed vs. healthy (°)	3.4	0.6	6.1	6.9	0.5	0.5	MES	1.4	3.5	NCR[a]
Pogonion-tragus-C7: myogenous vs. healthy (°)	2.0	−0.6	4.6	6.7	0.3	0.3	SES	1.3	3.4	NCR[a]
Tragus-C7-shoulder: mixed vs. healthy (°)	−4.9	−10.	0.4	13.1	−0.4	−0.4	SES	2.6	6.6	NCR[a]
Tragus-C7-shoulder: myogenous vs. healthy (°)	−1.2	−6.3	3.9	13.1	−0.1	−0.1	SES	2.6	6.6	NCR[a]

Criteria for scoring: When both the effect size (ES) and the mean difference between groups are higher than both MIDs (minimal important difference) then, scored as CR. If ES is moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small–moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small and one of the MID is accomplished, it is scored NCR. If both (ES and MID) are not accomplished or clinical criterion determines NCR, then it is scored NCR. Effect sizes are described according to Cohen, J: Small effect size: 0.20 (0–0.39): SES; medium effect size: 0.50 (0.4–0.79); MES; large effect size: $\geq$0.80: LEF.

[a] Based on clinical judgement: ES: effect size, MID: minimal important difference, NCR: not clinically relevant, PCR: potentially clinically relevant, CR: clinically relevant, SES: small effect size, MWS: moderate effect size, LEF: large effect size.

Table 2
Clinical relevance assessment of maximal cervical flexor strength, endurance of the flexor, and extensor cervical muscles outcome measures in patients with temporomandibular disorders when compared with healthy subjects.

Outcome	Mean difference	Confidence interval for mean difference		Pooled SD	Effect size (ES) or standardized mean difference	ES based on SD of control group	Interpretation ES	MID (0.2) = 0.2 × pooled SD	MID (0.5) = 0.5 × pooled SD	Final decision clinical relevance
		Lower	Upper							
Maximal cervical flexor strength										
Maximal strength: mixed TMD vs. healthy (N)	−3 N	−9.9	2.4	15.1	0.3	−0.3	SES	3.0	7.5	NCR[a]
Maximal strength: myogenous TMD vs. healthy (N)	−4.5 N	−10.3	1.4	15.0	0.3	−0.3	SES	3.0	7.5	NCR[a]
Endurance of the cervical flexor muscles										
Endurance of the neck flexor muscles at 25% mixed TMD vs. healthy (holding time in seconds)	−7.5 s	−12.4	−2.7	11.9	0.6	−0.6	MES	2.4	5.9	CR[a]
Endurance of the neck flexor muscles at 25% mixed TMD vs. myogenous (holding time in seconds)	−7.1 s	−11.8	−2.4	11.8	0.6	−0.6	MES	2.4	5.9	CR[a]
Endurance of the neck flexor muscles at 25% healthy vs. myogenous (holding time in seconds)	0.4 s	−4.2	5.0	11.8	0.04	0.04	SES	2.4	5.9	NCR[a]
Endurance of the cervical extensor muscles during NEMET										
NEMET endurance healthy vs. mixed TMD	207.0	39.8	374.2	408.0	0.5	0.5	MES	81.60	204.00	CR[a]
NEMET endurance healthy vs. myogenous TMD	211.0	51.6	370.5	408.0	0.5	0.5	MES	81.60	204.00	CR[a]

Criteria for scoring: when both the effect size (ES) and the mean difference between groups are higher than both MIDs (minimal important difference) then, scored as CR. If ES is moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small–moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small and one of the MID is accomplished, it is scored NCR. If both (ES and MID) are not accomplished or clinical criterion determines NCR, then it is scored NCR. Effect sizes are described according to Cohen, J: small effect size: 0.20 (0–0.39): SES; medium effect size: 0.50 (0.4–0.79): MES; large effect size: ≥0.80: LEF.

[a] Based on Clinical Judgement: ES: effect size; MID: minimal important difference; NCR: not clinically relevant; PCR: potentially clinically relevant; CR: clinically relevant; SES: small effect size; MWS: moderate effect size; LEF: large effect size.

strength between groups ranged between 3.73 and 4.45 N. The effect sizes of the differences were between 0.25 and 0.30 (small effect sizes). The MIDs in cervical flexor strength ranged between 3.0 and 7.50 N using 0.2 and 0.5 effect sizes respectively for the calculation. Because the calculated mean difference values of the difference between groups were lower than the MID values in addition to small effect sizes, a non-clinically relevant result was demonstrated (Table 2).

3.3. Endurance of the cervical flexor muscles

Subjects with mixed TMD had a statistically lower holding time than healthy subjects and subjects with myogenous TMD in the flexor muscle endurance test at 25% of the maximal voluntary contraction (MVC) condition. An average of almost 8 s difference in holding time between subjects with mixed TMD and healthy subjects and an average of 7 s difference between subjects with mixed TMD and those with myogenous TMD were found. The calculated effect sizes of the differences ranged between 0.60 and 0.63 (moderate effect sizes). The minimal important differences in holding time ranged between 2.36 and 5.94 s using 0.2 and 0.5 effect sizes respectively for the calculation. The calculated mean difference values of the difference between subjects with mixed and myogenous TMD and healthy subjects were higher than the MID values in addition to moderate effect sizes, demonstrating a clinically relevant result (Table 2).

3.4. Endurance of the cervical extensor muscles during NEMET

Subjects with mixed TMD and myogenous TMD had on average between 3.45 min (207 s) and 3.51 min (211 s) less holding time than healthy subjects respectively. The calculated effect sizes of the differences ranged between 0.50 and 0.52 (moderate effect sizes). The minimally important differences in holding time ranged between 1.36 min (81.6 s) and 3.4 min (204 s) using 0.2 and 0.5 effect sizes respectively for the calculation. The calculated mean difference values of the difference between subjects with mixed and myogenous TMD and healthy subjects were higher than the MID values, demonstrating a clinically relevant result (Table 2).

3.5. EMG activity of the cervical flexor muscles during the craniocervical flexion test (CCFT)

When performing the craniocervical flexion test, no statistically significant differences in electromyographic (EMG) activity in the sternocleidomastoid (SCM) muscles or the anterior scalene (AS) muscles in patients with mixed and myogenous TMD subjects were found when compared to healthy subjects ($p = 0.07$). However, when calculating the effect sizes of the differences in EMG activity of the SCM and AS muscles, moderate effect sizes ranging from 0.42 to 0.82 in many of the comparisons between subjects with TMD and healthy subjects were found. For a detailed description of the effect sizes as well as the MID for each level of pressure (22 mmHg,

Table 3
Clinical relevance assessment of the EMG activity of the cervical flexor muscles during the craniocervical flexion test (CCFT) in patients with temporomandibular disorders when compared with healthy subjects.

Outcome	Mean difference	Confidence interval for mean difference		Pooled SD	Effect size (ES) or standardized mean difference	ES based on SD of control group	Interpretation ES	MID (0.2) = 0.2 × pooled SD	MID (0.5) = 0.5 × pooled SD	Final decision clinical relevance
		Lower	Upper							
EMG activity of the cervical flexor muscles during the craniocervical flexion test (CCFT)										
EMG activity SCMR at 22 mmHg during the CCFT myogenous TMD vs. healthy (%MVC)	2.9	−0.8	6.5	9.2	0.3	0.4	SES to MES	1.8	4.6	NCR[a]
EMG activity SCMR at 22 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.4	1.7	9.1	9.2	0.6	0.7	MES	1.8	4.6	CR[a]
EMG activity SCMR at 24 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.1	−0.9	7.0	10	0.3	0.4	SES to MES	2.0	5.0	NCR[a]
EMG activity SCMR at 24 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.9	1.8	9.9	10	0.6	0.7	MES	2.0	5.0	CR[a]
EMG activity SCMR at 26 mmHg during the CCFT myogenous vs. healthy (%MVC)	1.7	−3.3	6.7	12.7	0.1	0.2	SES	2.5	6.3	NCR[a]
EMG activity SCMR at 26 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	4.2	−0.8	9.2	12.3	0.3	0.4	SES to MES	2.5	6.2	NCR[a]
EMG activity SCMR at 28 mmHg during the CCFT myogenous vs. healthy (%MVC)	1.9	−3.4	7.2	13.4	0.1	0.2	SES	2.7	6.7	NCR[a]
EMG activity SCMR at 28 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.9	0.8	11.1	12.8	0.5	0.5	MES	2.6	6.4	PCR[a]
EMG activity SCMR at 30 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.8	−2.5	10.1	16.0	0.2	0.3	SES	3.2	8.0	NCR[a]
EMG activity SCMR at 30 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	6.3	0.7	12.0	13.9	0.5	0.5	MES	2.8	7.0	CR[a]
EMG activity SCML at 22 mmHg during the CCFT myogenous vs. healthy (%MVC)	5.1	0.7	9.5	11.2	0.5	0.7	MES	2.2	5.6	PCR[a]
EMG activity SCML at 22 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.8	2.1	9.5	9.2	0.6	0.8	MES to LES	1.8	4.6	CR[a]
EMG activity SCML at 24 mmHg during the CCFT myogenous vs. healthy (%MVC)	4.9	0.8	9.0	10.3	0.5	0.7	MES	2.1	5.2	PCR[a]
EMG activity SCML at 24 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	6.5	2.5	10.6	10.0	0.7	0.9	MED to LES	2.0	5.0	CR[a]
EMG activity SCML at 26 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.9	−0.9	8.8	12.2	0.3	0.4	SES to MES	2.4	6.1	PCR[a]
EMG activity SCML at 26 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	4.6	0.3	9.0	10.8	0.4	0.5	MES	2.2	5.4	PCR[a]
EMG activity SCML at 28 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.0	−2.5	8.5	13.8	0.2	0.3	SES	2.8	6.9	NCR[a]
EMG activity Av SCML at 28 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	4.6	−0.2	9.4	11.7	0.4	0.4	MES	2.3	5.9	PCR[a]
EMG activity SCML at 30 mmHg during the CCFT myogenous vs. healthy (%MVC)	5.1	−1.3	11.5	16.2	0.3	0.4	SES to MES	3.2	8.1	PCR[a]
EMG activity SCML at 30 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.2	0.1	10.3	12.7	0.4	0.4	MES	2.5	6.3	PCR[a]
EMG activity ASR at 22 mmHg during the CCFT myogenous vs. healthy (%MVC)	6.4	0.5	12.3	14.9	0.4	0.6	MES	3.0	7.5	PCR[a]
EMG activity ASR_22 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	3.9	−1.0	8.7	11.9	0.3	0.4	SES to MES	2.4	5.9	NCR[a]
EMG activity ASR at 24 mmHg during the CCFT myogenous vs. healthy (%MVC)	5.6	−0.2	11.4	14.6	0.4	0.5	SES to MES	2.9	7.3	PCR[a]
EMG activity ASR at 24 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	4.9	−0.7	10.4	13.7	0.4	0.4	SES to MES	2.7	6.9	PCR[a]
EMG activity ASR at 26 mmHg during the CCFT myogenous vs. healthy (%MVC)	5.8	−0.8	12.4	16.6	0.4	0.4	SES to MES	3.3	8.3	PCR[a]

EMG activity ASR at 26 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.2	−1.9	12.4	17.7	0.3	0.4	SES	3.5	8.8	NCR[a]
EMG activity ASR at 28 mmHg during the CCFT myogenous vs. healthy (%MVC)	6.6	−1.5	14.6	20.3	0.3	0.4	SES to MES	4.1	10.2	PCR[a]
EMG activity ASR at 28 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	6.6	−0.5	13.7	17.6	0.4	0.4	SES to MES	3.5	8.8	PCR[a]
EMG activity ASR at 30 mmHg during the CCFT myogenous vs. healthy (%MVC)	12.1	1.0	23.1	27.9	0.4	0.7	MES	5.6	14.0	PCR[a]
EMG activity ASR at 30 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	8.2	0.2	16.3	19.9	0.4	0.5	MES	4.0	10.0	PCR[a]
EMG activity ASL at 22 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.8	−1.2	8.7	12.5	0.3	0.4	SES to MES	2.5	6.3	NCR[a]
EMG activity ASL at 22 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	4.4	−1.2	9.9	13.7	0.3	0.4	SES to MES	2.7	6.8	PCR[a]
EMG activity ASL at 24 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.0	−2.1	8.1	12.8	0.2	0.3	SES	2.6	6.4	NCR[a]
EMG activity ASL at 24 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.4	−1.1	11.8	15.9	0.3	0.5	SES to MES	3.2	7.9	PCR[a]
EMG activity ASL at 26 mmHg during the CCFT myogenous vs. healthy (%MVC)	1.6	−4.4	7.6	15.2	0.1	0.1	SES	3.0	7.6	NCR[a]
EMG activity ASL at 26 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	3.5	−3.7	10.8	17.8	0.2	0.3	SES	3.6	8.9	NCR[a]
EMG activity ASL at 28 mmHg during the CCFT myogenous vs. healthy (%MVC)	3.3	−5.9	12.6	23.5	0.1	0.2	SES	4.7	11.7	NCR[a]
EMG activity ASL at 28 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	5.7	−2.1	13.5	19.3	0.3	0.4	SES to MES	3.9	9.7	PCR[a]
EMG activity ASL at 30 mmHg during the CCFT myogenous vs. healthy (%MVC)	6.5	−3.2	16.2	24.4	0.3	0.5	SES to MES	4.9	12.2	PCR[a]
EMG activity ASL at 30 mmHg during the CCFT mixed TMD vs. healthy (%MVC)	6.3	−1.5	14.1	19.2	0.3	0.4	SES to MES	3.8	9.6	PCR[a]

Criteria for scoring: When both the effect size (ES) and the mean difference between groups are higher than both MIDs (minimal important difference) then, scored as CR. If ES is moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small−moderate and one of the MIDs is accomplished, it is scored PCR. If ES is small and one of the MID is accomplished, it is scored NCR. If both (ES and MID) are not accomplished or clinical criterion determines NCR, then scored NCR. Effects sizes are described according to Cohen, J: Small effect size: 0.20 (0−0.39): SES; medium effect size: 0.50 (0.4−0.79): MES; large effect size: ≥0.80: LEF.

[a] Based on clinical judgement: ES: effect size, MID: minimal important difference, NCR: not clinically relevant, PCR: potentially clinically relevant, CR: clinically relevant, SES: small effect size, MWS: moderate effect size, LEF: large effect size.

24 mmHg, 26 mmHg, 28 mmHg, 30 mmHg) and each muscle (SCM and AS) see Table 3.

4. Discussion

This study shows an example of how to evaluate the clinical relevance of research results using data obtained from a cross-sectional study comparing several outcomes used for evaluating neck musculoskeletal functioning in patients with TMD when compared with healthy subjects. The results of this study show that it is possible to have statistical significance without having clinical relevance, to have both statistical significance and clinical relevance, to have clinical relevance without having statistical significance, or to have neither statistical significance nor clinical relevance.

The use of different methods to determine clinical relevance can help clinicians to interpret research results. The methods presented here are simple and do not require complex calculations, so can be easily applied by any clinician or researcher. The use of different methods will depend on several factors. When data has been already collected, distribution-based methods (i.e. effect size, MID) are the methods of choice (Musselman, 2007). However, if the study is being designed, anchor-based methods (i.e. Global Rating Scale), involving the clients perspectives, should be implemented (Musselman, 2007). In this example, we used the distribution-based methods because they can be calculated retrospectively once the data has been collected (as it is the case of this study) and data regarding means and standard deviations for the outcomes by groups were available.

4.1. Statistical significance and questionable clinical relevance

In our example, the difference between myogenous and healthy controls in the craniocervical posture (measured using the eye-tragus-horizontal angle) was 2.6°. According to the statistical test, this result was deemed to be statistical significant. This value, although higher that the MID and having an effect size of 0.48, was questionably clinically relevant based on our "clinical judgement" since it is very unlikely that clinicians in routine examination would be able to detect such a small difference using conventional postural screening even using a newly developed-sophisticated digital inclinometer. The standard error of measurement of this device ranged between 1.6° and 2.6°, which includes the mean difference obtained in our study (Prushansky et al., 2010). In addition, clinicians would not normally change their treatment intervention and progression based on that small difference which could be just an "error". In this case, clinical judgement, based on clinical knowledge and previous information of the outcome will take precedence over the calculated methods for determining clinical relevance of these results. This decision regarding clinical relevance is in line with Kirk (Kirk, 1996) and Callahan's (Callahan and Reio, 2006) suggestions regarding interpretation of effect size values. According to these authors, these values are used only as guidelines and cannot be "sanctified". They need to be used in combination with previous knowledge and standards of clinical relevance and also using clinical reasoning.

4.2. Statistical significance and clinical relevance

In the case of endurance of the cervical flexor and extensor muscles measured through the holding time in seconds and minutes respectively, the difference in holding time between groups was considered both statistically significant and clinically significant. Because normative values for holding time for both group of muscles are not determined, and also because of the large variability found in different research investigating holding times for cervical flexors (Barber, 1994; Grimmer, 1994; Blizzard et al., 2000; Olson et al., 2006) as well as for cervical extensors (Lee et al., 2004; Lee et al., 2005; Peolsson et al., 2007), statistical significance is insufficient by itself for the clinician to know if these results are indeed of importance. Thus, the evaluation of clinical relevance needs to be performed in order to determine whether these results could be of value for the clinicians. The calculated ES for these outcomes ranged between 0.63 and 0.51 for flexors and extensors muscles respectively. These effect sizes are considered moderate effects, which in simple terms, means that the effect size (difference between groups) is "visible to the naked eye of a careful observer" (Cohen, 1988). In addition the mean differences between groups (i.e. healthy and TMD groups) obtained in these outcomes were higher than the calculated MIDs. Therefore, these findings should be taken into consideration when evaluating and treating subjects with TMD. People with TMD might have impairment in the endurance of the flexor and extensor muscles and might benefit from treatment of this impairment.

4.3. Not statistically significant but clinically relevant

Another case would be when statistical significance is not attained but clinical relevance is. In our study, it was found that subjects with TMD had no statistically significant differences in EMG activity of the superficial cervical muscles (SCM and AS) when compared to healthy subjects ($p = 0.07$), although important effects sizes reflecting a clinically important difference between the two groups were found. In this case, the standards for determining clinical relevance based on the outcome (EMG activity) are not known. Although, there is no research establishing a cut off of EMG activity (% MVC) to be considered clinically important when comparing the EMG activity of different groups, it has been seen that electromyographic activities as low as 2–5% of the MVC can be related to pain in neck–shoulder areas (Jonsson, 1988; Veiersted et al., 1990; Jensen et al., 1993). So, in this case one does have some background information to evaluate the importance of the findings, although this is very limited. In this case, we can base our decision on the calculated effect size. Thus, standardized effect sizes and MID could serve as an index to guide clinicians in the relevance of the findings. It could be said that in absence of knowledge and guidelines to determine the clinical relevance of a certain outcome, calculation of the clinical relevance, based on the distribution methods could be an option; however, they cannot be considered as absolute values of clinical relevance in all conditions.

In our example, although statistical significance was not attained, ES and MID values indicated that an important difference in electromyographic activity between subjects with TMD and healthy subjects in several conditions existed. Although variability of the electromyographic activity was high, subjects with TMD had a strong tendency to have increased EMG activity of the superficial cervical muscles when compared with healthy subjects. This could indicate a different strategy by the subjects with TMD for activating the cervical muscles to stabilize the craniocervical system when compared with pain free subjects.

4.4. Not statistical significant and questionable clinical relevance

There are cases in which statistical significance and clinical relevance both are "not significant". For example, in our study, maximal cervical flexor strength was neither statistically significant nor clinically relevant. Thus, when both methods agree on significance, it is easier for clinicians to determine the implications of these results and to decide whether or not they will use this information in clinical practice. Thus, we can say that subjects with TMD presented with differences that were not statistically nor

clinically important for maximal cervical flexor strength when compared with healthy individuals in this study.

4.5. Limitations

It is important to highlight that the evaluation of clinical relevance and its implications are applicable for the group of subjects who participated in this study under the protocols used. They could potentially be applied to subjects with TMD having similar clinical characteristics as the subjects who participated in this study and outcomes could be measured in a similar way. This limitation should be taken into consideration when attempting to extrapolate these results.

A high prevalence of neck pain existed in the TMD population analyzed in this study. Almost 88% of the subjects with TMD had self-reported neck pain. Only 13 subjects (7 subjects with myogenous TMD and 6 subjects with mixed TMD) from the entire group of TMD patients were not affected by neck pain at all. The same proportion of patients with isolated TMD has been found in a previous study (Lobbezoo et al., 2004). This finding attests for a high prevalence of neck pain in the TMD population and very low probability of finding subjects with isolated TMD. In fact, according to a study performed by Stiesch-Scholz et al. (Stiesch-Scholz et al., 2003), asymptomatic functional disorders of the cervical spine occurred more frequently in patients with internal derangement of the TMJ than in a control group. Thus, cervical spinal disorders (CSD) could be present in the TMD population even if they are not symptomatic. Thus, CSD could be a condition that is part of TMD. Thus, we see neck pain as an "intervening variable" as so called by Wunsch (2007). Since Cervical Spinal Disorders (CSD) are so common in patients with TMD, we preferred to analyze the data all together to provide clinicians with the situation they would most commonly see in clinical practice. If a patient with TMD came to their clinic, there is a high *probability* that the patient would present with CSD. These results indicate that if patients with TMD have neck pain and neck disability, the treatment may need to focus on both areas since the improvement of one could have an influence on the other. Thus, this manuscript highlights the fact that assessment and treatment for patients with TMD needs to consider neck involvement in TMD patients.

4.6. Issues when evaluating clinical relevance

When evaluating clinical relevance based on distribution-based methods, we realized that sometimes the methods do not agree in the results. Thus, we developed some criteria to determine the clinical relevance of the results. It is desirable that researchers and clinicians make an *a priori* criterion to determine clinical relevance of the results and use multiple methods to evaluate the consistence of the results (i.e. effect size, MID, standard error of the measurements). In addition, they can use background information regarding criteria of clinical relevance from previous studies (if available) to make the final decision.

In rehabilitation disciplines, the client's perspective is highly valued. Thus, when possible, clinical researchers should evaluate clinical significance based on anchor methods (i.e. Global Rating Scale), in addition to distribution-based methods. In this way, the input of the participants can be taken into account.

Clinical relevance is a complex construct and has to be analyzed from different points of view such as type of pathology, clients' perspectives of improvement, as well as societal perspectives regarding impact on public safety, health care policy and cost. Thus, clinical relevance assessment methods can help guide decisions regarding the clinical importance of results but they need to be complemented with clinical reasoning as well as clinical experience or clients' perspectives when available in order to make a final decision (Callahan and Reio, 2006).

The evaluation of the clinical relevance has been advocated for a long time but still has not been implemented or reported in most research reports (Faulkner et al., 2008). Clinical researchers have the responsibility of assessing and reporting the clinical relevance of their research results in addition to the analysis of statistical significance to simplify the transfer of knowledge from research into practice. In this way, knowledge translation might be improved which in turn will result in a better quality of care.

5. Conclusion

The evaluation of clinical relevance in clinical research is crucial to simplify the transfer of knowledge from research into practice. Clinicians and researchers need to be aware of the importance of the research results and should abandon the only simplistic approach of statistical significance interpretation. This paper encourages researchers to assess and present the clinical relevance of their research results in addition to the statistical significance analysis. In addition, editors of scientific journals should encourage authors to report the clinical relevance of their results. In this way, knowledge translation might be improved.

Acknowledgements

Alberta Provincial CIHR Training Program in Bone and Joint Health, Canadian Institutes of Health Research, Government of Chile (MECESUP Program), University Catholic of Maule, Physiotherapy Foundation of Canada through an Alberta Research Award and the University of Alberta. Authors would like to thank Martha Funabashi, Larissa Costa, and Anelise Silveira for their constructive feedback.

References

Armijo-Olivo S. Relationship between cervical musculoskletal impairments and temporomandibular disorders: clinical and electromyographic variables, faculty of rehabilitation medicine. Edmonton: University of Alberta; 2010. p. 488.

Armijo-Olivo S, Fuentes JP, da Costa BR, Major PW, Warren S, Thie NMR, et al. Reduced endurance of the cervical flexor muscles in patients with concurrent temporomandibular disorders and neck disability. Manual Therapy 2010a; 15(6):586–92.

Armijo-Olivo S, Rappoport K, Fuentes J, Gadotti IC, Major P, Warren S, et al. Head and cervical posture in patients with temporomandibular disorders (TMD). Journal of Orofacial Pain, in press-a.

Armijo-Olivo S, Silvestre R, Fuentes J, da Costa BR, Gadotti IC, Warren S, et al. Electromyographic activity of the cervical flexor muscles in patients with temporomandibular disorders while executing the craniocervical flexion test (CCFT): a cross sectional study. Physical Therapy, in press-b.

Armijo-Olivo S, Silvestre R, Fuentes J, da Costa BR, Warren S, Major P, et al. Patients with temporomandibular disorders have increase fatigability of the cervical extensor muscles. The Clinical Journal of Pain, in press-c.

Armijo-Olivo SA, Fuentes J, Major PW, Warren S, Thie NMR, Magee DJ. The association between neck disability and jaw disability. Journal of Oral Rehabilitation 2010b;37(9):670–9.

Armijo-Olivo SL, Fuentes JP, Major PW, Warren S, Thie NM, Magee DJ. Is maximal strength of the cervical flexor muscles reduced in patients with temporomandibular disorders? Archives of Physical Medicine and Rehabilitation 2010c; 91(8):1236–42.

Barber A. Upper cervical spine flexor muscles: age related performance in asymptomatic women. Australian Journal of Physiotherapy 1994;40(3):167–71.

Blizzard L, Grimmer KA, Dwyer T. Validity of a measure of the frequency of headaches with overt neck involvement, and reliability of measurement of cervical spine anthropometric and muscle performance factors. Archives of Physical Medicine and Rehabilitation 2000;81(9):1204–10.

Callahan JL, Reio Jr TG. Making subjective judgments in quantitative studies: the importance of using effect sizes and confidence intervals. Human Resource Development Quarterly 2006;17(2):159–73.

Carnine D. Bridging the research-to-practice gap. Exceptional Children 1997;63(4): 513–21.

Cesar GM, Tosato JP, Biasotto-Gonzalez DA. Correlation between occlusion and cervical posture in patients with bruxism. Compendium of Continuing Education in Dentistry 2006;27(8):463–6 [Jamesburg, N.J.: 1995].

Cohen J. The concepts of power analysis. In: Cohen J, editor. Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Academic Press, Inc; 1988. p. 1–17 [chapter 1].
Cooke MS, Wei SH. The reproducibility of natural head posture: a methodological study. American Journal of Orthodontics and Dentofacial Orthopedics. Official publication of the American Association Of Orthodontists, its constituent societies, and the American Board Of Orthodontics. 1988;93(4):280–8.
Falla D, Jull G, Dall'Alba P, Rainoldi A, Merletti R. An electromyographic analysis of the deep cervical flexor muscles in performance of craniocervical flexion. Physical Therapy 2003;83(10):899–906.
Faulkner C, Fidler F, Cumming G. The value of RCT evidence depends on the quality of statistical analysis. Behaviour Research and Therapy 2008;46(2):270–81.
Gadotti IC, Berzin F, Biasotto-Gonzalez D. Preliminary rapport on head posture and muscle activity in subjects with class I and II. Journal of Oral Rehabilitation 2005;32(11):794–9.
Grimmer K. Measuring the endurance capacity of the cervical short flexor muscle group. Australian Journal of Physiotherapy 1994;40(4):251–4.
Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Aaronson N, et al. Methods to explain the clinical significance of health status measures. Mayo Clinic Proceedings 2002;77(4):371–83.
Jensen BR, Schibye B, Sogaard K, Simonsen EB, Sjogaard G. Shoulder muscle load and muscle fatigue among industrial sewing-machine operators. European Journal of Applied Physiology and Occupational Physiology 1993;67(5):467–75.
Jonsson B. The static load component in muscle work. European Journal of Applied Physiology and Occupational Physiology 1988;57(3):305–10.
Kirk RE. Practical significance: a concept whose time has come. Educational and Psychological Measurement 1996;56(5):746–59.
Kirk RE. Effect magnitude: a different focus. Journal of Statistical Planning and Inference 2007;137(5):1634–46.
Lee H, Nicholson LL, Adams RD. Cervical range of motion associations with subclinical neck pain. Spine 2004;29(1):33–40.
Lee H, Nicholson LL, Adams RD. Neck muscle endurance, self-report, and range of motion data from subjects with treated and untreated neck pain. Journal of Manipulative and Physiological Therapeutics 2005;28(1):25.
Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. Journal of Clinical Epidemiology 2007;60(5):448–55.
Lobbezoo F, Visscher CM, Naeije M. Impaired health status, sleep disorders, and pain in the craniomandibular and cervical spinal regions. European Journal of Pain 2004;8(1):23–30.
McNeely M, Warren S. Value of confidence intervals in determining clinical significance. Physiotherapy Canada 2006;58:205–11.
Millis SR. Emerging standards in statistical practice: implications for clinical trials in rehabilitation medicine. American Journal of Physical Medicine and Rehabilitation 2003;82(10 Suppl.).
Musselman KE. Clinical significance testing in rehabilitation research: what, why, and how? Physical Therapy Reviews 2007;12(4):287–96.
Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: history, application, and current practice. Clinical Psychology Review 2001;21(3):421–46.
Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemporary Educational Psychology 2000; 25(3):241–86.
Olson LE, Millar AL, Dunker J, Hicks J, Glanz D. Reliability of a clinical test for deep cervical flexor endurance. Journal of Manipulative and Physiological Therapeutics 2006;29(2):134–8.
Peolsson A, Almkvist C, Dahlberg C, Lindqvist S, Pettersson S. Age- and sex-specific reference values of a test of neck muscle endurance. Journal of Manipulative and Physiological Therapeutics 2007;30(3):171–7.
Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. Journal of the American Geriatrics Society 2006;54(5):743–9.
Portney L, Watkins M. In: Mehalik C, editor. Power and sample size. Foundations of Clinical Research, New Jersey: Prentice Hall Health; 2000. p. 705–30.
Prushansky T, Deryi O, Jabarreen B. Reproducibility and validity of digital inclinometry for measuring cervical range of motion in normal subjects. Physiotherapy Research International: The Journal for Researchers and Clinicians in Physical Therapy 2010;15(1):42–8.
Sandoval P, Henriquez J, Fuentes R, Cabezas G, Roldan R. Cervical curve: cephalometric study in rest clinical position. Revista Medica de Chile 1999;127:547–55.
Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. COPD: Journal of Chronic Obstructive Pulmonary Disease 2005;2(1):57–62.
Sterne JAC, Smith GD. Sifting the evidence – what's wrong with significance tests? British Medical Journal 2001;322(7280):226–31.
Stiesch-Scholz M, Fink M, Tschernitschek H. Comorbidity of internal derangement of the temporomandibular joint and silent dysfunction of the cervical spine. Journal of Oral Rehabilitation 2003;30(4):386–91.
Thompson B. Improving research clarity and usefulness with effect size indices as supplements to statistical significance tests. Exceptional Children 1999;65(3): 329–37.
Veiersted KB, Westergaard RH, Andersen P. Pattern of muscle activity during stereotyped work and its relation to muscle pain. International Archives of Occupational and Environmental Health 1990;62(1):31–41.
Wunsch G. Confounding and control. Demographic Research 2007;16:97–120.