

## University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

**UNIVERSITY OF SOUTHAMPTON**

**FACULTY OF MEDICINE, HEALTH & LIFE SCIENCES**

**School of Biological Sciences**

**Investigation of Head Repositioning Accuracy as a Measure of  
Cervicocephalic Kinaesthetic Sensibility in Patients with Chronic Neck Pain**

**by**

**George D. W. Rix BSc (Chiro), DC, FEAC, FHEA.**

**Thesis for the degree of Doctor of Philosophy**

**October 2008**

**ABSTRACT**

**FACULTY OF MEDICINE, HEALTH & LIFE SCIENCES**

**SCHOOL OF BIOLOGICAL SCIENCES**

**Doctor of Philosophy**

**INVESTIGATION OF HEAD REPOSITIONING ACCURACY AS A MEASURE OF CERVICOCEPHALIC KINAESTHETIC SENSIBILITY IN PATIENTS WITH CHRONIC NECK PAIN**

**by George Douglas William Rix**

The majority of people can expect to experience neck pain in their lifetime and some will go on to develop prolonged or repetitive episodes of neck pain or related symptoms. These persistent complaints have become a major cause of disability around the world.

Although chronic 'mechanical' neck pain can be defined in clinical terms, the underlying pathology remains unclear. Research has failed to demonstrate a consistent relationship between the presence of neck pain and pathology such as degenerative changes. As such, there has been an increasing interest in altered neuro-muscular-articular function in the pathogenesis of neck pain. Over the last 17 years, the role of cervical proprioceptive (mechanoreceptive) dysfunction in the perpetuation of chronic neck pain has received increasing attention from researchers and clinicians. This is commonly referred to as cervicocephalic kinaesthetic sensibility (KS).

Cervicocephalic KS has generally been studied utilising head repositioning accuracy (HRA) tasks. At the beginning of this study only seven reports had been published in the area. Following review of this literature, several focused areas of interest were apparent for further study - 1) comparison of KS in various patient subgroups (e.g., insidious onset vs 'whiplash'); 2) development of more sophisticated methods of measuring head movement and repositioning errors; 3) establishing the characteristics of the tests such as method agreement and reliability; 4) comparison of the discriminative value of repositioning tasks to both subjective 'straight ahead' (SSA) and to non-neutral set points within the cervical range of motion (non-straight ahead or nSA).

*Investigation 1* used a laser pointer method to study head repositioning errors in patients with chronic neck pain of insidious onset. The results suggested that these patients with chronic neck pain show little evidence of impaired cervicocephalic KS, when measured as HRA-SSA. The study also served to highlight several difficulties with the laser pointer method of measurement and the relatively poor knowledge of HRA in healthy subjects. Most previous studies used the mean of 10 repetitions for the measurement but more recent studies utilised fewer repetitions. Although the laser pointer method is simple, inexpensive and easy to use, the method involves a degree of experimenter bias and inaccuracy. It also does not lend itself to concurrent evaluation of variables such as range of motion and speed of head movement.

*Investigations 2 & 3* focused on the development of the testing method and the introduction of the Zebris CMS 70P ultrasound system for the recording of HRA-SSA, namely; method agreement between a 5 and 10 repetition measuring protocol; method agreement between the laser pointer and Zebris system; the intra/inter-examiner reliability of measurement methods.

The Zebris system results suggested that the two methods of measurement agree sufficiently well for the 5 repetition method to replace the 10 repetition method to obtain a mean HRA score and that both could be used interchangeably. Further results suggested that the Zebris and laser pointer methods do not agree sufficiently well to be used interchangeably. The test-retest reliability was comparable between both methods suggesting that from this perspective, either could be used for measuring HRA-SSA. The inter-rater test-retest reliability was comparable to the test-retest reliability suggesting that trained examiners could be interchanged when carrying out repeated measurements.

*Investigation 4*, the final study in this thesis, investigated HRA with the Zebris system using the 5 repetition protocol in two groups of chronic neck pain patients; insidious onset and neck pain from a 'whiplash' injury. The results suggested that patients with chronic neck pain of both insidious onset and from a 'whiplash' injury show little evidence of impaired cervicocephalic KS when measured using HRA-SSA and nSA tests. These results conflict with previous studies

Despite numerous investigations over the last 17 years, a test that can be routinely applied in the clinical setting for the purposes of diagnosis and treatment monitoring has not been established. Studies to date suggest that an active HRA test to SSA/NHP that is established by the patient may have the greatest discriminative value. Although HRA testing shows some promise in identifying deficits in 'whiplash' patients, it is unlikely that the KS tests identify specific subgroups of chronic neck pain patients. There have been several contradictory studies which have shown considerable overlap between patient and healthy groups. It is also unlikely that HRA tests represent a unique test of cervical proprioceptive function (peripheral or central integration) and therefore provide a test exclusive to neck disorders.

# List of Contents

	<b>Abstract</b>	<b>i</b>
	<b>List of Contents</b>	<b>ii</b>
	<b>List of Figures and Illustrations</b>	<b>iv</b>
	<b>List of Tables</b>	<b>viii</b>
	<b>Declaration of Authorship</b>	<b>xii</b>
	<b>Acknowledgements</b>	<b>xiii</b>
	<b>Abbreviations</b>	<b>xiv</b>
	<b>Dedication</b>	<b>xvi</b>
<b>Chapter 1</b>	<b>General Introduction and Aims of the Research</b>	<b>1</b>
	1.1 History, Background & Statement of the Problem	1
	1.2 Functional Deficits – Cervical Spine Proprioception	6
	1.3 Aims and Objectives of the Research	15
<b>Chapter 2</b>	<b>Head Repositioning Accuracy using the Laser Pointer Method in Patients with Chronic Insidious Onset Neck Pain</b>	<b>33</b>
	2.1 Introduction	33
	2.2 Methods	35
	2.3 Results	43
	2.4 Discussion	54
	2.5 Conclusion	63
<b>Chapter 3</b>	<b>Head Repositioning Accuracy: A Critical Review of ‘Reliability’ Characteristics and Methods of Statistical Analysis</b>	<b>64</b>
	3.1 Introduction	64
	3.2 The Development and Characteristics of Clinical Tests	65
	3.3 Statistical Approaches for the Quantification of Agreement	72
<b>Chapter 4</b>	<b>Development of Head Repositioning Accuracy Testing using the Zebris CMS 70P System: A Study of Method Agreement</b>	<b>120</b>
	4.1 Introduction	120
	4.2 Methods	122
	4.3 Results	132
	4.4 Discussion	147
	4.5 Conclusion	152

<b>Chapter 5</b>	<b>Inter &amp; Intra Examiner ‘Reliability’ and Method Agreement of the Zebris and Laser Pointer Methods for Measuring Head Repositioning Accuracy</b>	<b>153</b>
	5.1 Introduction	153
	5.2 Methods	157
	5.3 Results	163
	5.4 Discussion	194
	5.5 Conclusion	200
<b>Chapter 6</b>	<b>Head Repositioning Accuracy In Patients with Chronic, Traumatic and Insidious Onset Neck Pain: A Study of Three Repositioning Tests.</b>	<b>201</b>
	6.1 Introduction	201
	6.2 Methods	204
	6.3 Results	208
	6.4 Discussion	225
	6.5 Conclusion	234
<b>Chapter 7</b>	<b>General Discussion</b>	<b>235</b>
<b>Chapter 8</b>	<b>Overall Conclusions</b>	<b>242</b>
	<b>References</b>	<b>244</b>
	<b>Appendix I</b>	<b>252</b>
	<b>Appendix II</b>	<b>254</b>
	<b>Appendix III</b>	<b>255</b>
	<b>Appendix IV</b>	<b>262</b>
	<b>Appendix V</b>	<b>277</b>

## List of Figures and Illustrations

<b>Figure 1.</b> Graphical representation of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and its Associated Disorders Conceptual Model for the onset, course, and care of neck pain. Adapted from Guzman et al (2008). .....	5
<b>Figure 2.</b> Simplified schematic of the neural subsystems, reflexes and connections for head and neck position sense and motor control (Armstrong, McNair and Taylor, 2008).....	7
<b>Figure 3.</b> Cervical range of motion device (CROM) .....	37
<b>Figure 4.</b> Experimental apparatus and testing procedure used for evaluation of head repositioning accuracy. ....	38
<b>Figure 5.</b> HRA data collection on target and initial analysis.....	41
<b>Figure 6.</b> A) Linear (displacement) and B) Cartesian (angular) axes.....	42
<b>Figure 7A-F.</b> Box plots of active cervical range of motion data for patient and control groups in each of the movement directions. ....	45
<b>Figure 8.</b> Scatterplots showing the head repositioning accuracy (HRA) for the neck pain patients and control subjects.....	47
<b>Figure 9 A-D.</b> Box plots of absolute global (R), horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) for the neck pain patients and control subjects.....	48
<b>Figure 10.</b> Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients and control subjects.....	52
<b>Figure 11.</b> Diagrammatic representation of BSI definitions for the accuracy of a measurement method for measurements on a continuous scale that yield a single value .....	70
<b>Figure 12.</b> Suggested terminology for common 'Reliability' studies using BSI framework. ....	71
<b>Figure 13.</b> Graphical representation of systematic biases and error types between two measurements using scatterplots. ....	74
<b>Figure 14.</b> Scatterplots of artificial measurements. A) XY data set – random error; B) XY <sub>1</sub> data set – random error and fixed bias. ....	78
<b>Figure 15.</b> Scatterplots of artificial measurements. A) XY <sub>2</sub> data set – random error and proportional bias; B) XY <sub>3</sub> data set – random error and fixed bias.....	80
<b>Figure 16.</b> Scatterplots of artificial measurements. A) X <sub>1</sub> Y <sub>4</sub> data set – increased random error; B) XY <sub>5</sub> data set – heteroscedasticity; C) & D) residuals against predicted values (regression standardised) for data sets X <sub>1</sub> Y <sub>4</sub> and XY <sub>5</sub> respectively.....	82
<b>Figure 17.</b> Scatterplots of artificial measurements. A) XY data set – low range ; B) XY data set – high range; C) X <sub>1</sub> Y <sub>4</sub> data set – low range ; D) X <sub>1</sub> Y <sub>4</sub> data set – high range. ....	83
<b>Figure 18.</b> Flow chart for selecting an appropriate intraclass correlation coefficient (ICC). Adapted from McGraw and Wong (1996) .....	89
<b>Figure 19.</b> Graphical representation of agreement definitions in various correlation coefficients. ICC = Intraclass correlation coefficient (adapted from Schuck, 2004).....	90
<b>Figure 20.</b> Hypothetical scatterplot showing the output components of regression analysis that may be used for the assessment of agreement. ....	97

<b>Figure 21.</b> A) – Residuals types used with variations of linear regression analysis. Solid line – line of best fit; $\Delta Y$ , vertical distance of x,y point from line; $\Delta X$ , horizontal distance of x,y point from line; $\Delta P$ perpendicular distance of x,y point from line. B) – Hypothetical example showing the variation in regression line (dotted line) with Ordinary Least Squares (OLS) of Y and OLS of X regression techniques. ....	99
<b>Figure 22.</b> A) – Scatterplot for data set $X_1Y_4$ ; solid line – line of equality (true relationship measurements; B) Ordinary Least Squares (OLS) of Y regression analysis for data set $X_1Y_4$ showing the attenuation of regression line with this technique (dotted line).....	100
<b>Figure 23.</b> A) - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set. B) – Graph representing the mean of the paired differences compared with zero (no difference).....	106
<b>Figure 24.</b> A - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set. B – Graph representing the mean of the paired differences compared with zero (no difference).....	107
<b>Figure 25.</b> A - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set (dotted & solid lines represent difference in a –ve or +ve direction). B – Graph representing the mean of the paired differences compared with zero (no difference).. ....	108
<b>Figure 26.</b> Bland and Altman distribution plot.....	112
<b>Figure 27.</b> Graphical comparison of error and bias types using scatterplots (A,C,E,G) and Bland and Altman distribution plots (B,D,F,H).....	113
<b>Figure 28.</b> Diagrammatic representation of method agreement study design and data analysis.. ....	122
<b>Figure 29.</b> Zebris CMS 70P system.. ....	124
<b>Figure 30.</b> Earphone system and attachment to the headpiece. ....	126
<b>Figure 31.</b> General view of the LabVIEW™ application measurement analysis window with cursors removed.....	129
<b>Figure 32.</b> First HRA-SSA movement repetition for a $RR \Rightarrow 0$ trial. The coloured, labelled (e.g, CX or C2) vertical lines are the cursors used to obtain the HRA, ROM & velocity measurements. Cursors: CX – CY, reference zero for HRA; C1, start of movement away from reference zero; C2-C3, peak ROM; C4, initial cessation of trial movement; C5-C6, SSA measurement.....	130
<b>Figure 33.</b> Scatter plots showing the individual subject head repositioning accuracy to subjective straight ahead (HRA-SSA ) in the two horizontal rotation directions (L & RR) for each of the trials at the test and retest sessions.. ....	133
<b>Figure 34.</b> Box plots of absolute horizontal rotation ( $\theta Y$ ) and vertical rotation ( $\theta X$ ) repositioning error (degrees) for each of the trials at the test and retest sessions.. ....	134
<b>Figure 35.</b> Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) in healthy subjects (n=13). ....	136
<b>Figure 36.</b> Schematic overview of LR & RR data sets for the Bland & Altman method agreement (5 vs 10 reps trials) that exhibit <i>Heteroscedasticity &amp;/or Proportional Bias</i> .....	139
<b>Figure 37.</b> Plots of the difference against mean for the absolute 5 and 10 repetitions X-axis head repositioning accuracy (HRA) at the first testing session (A) and with replicate measurements (B).....	140

<b>Figure 38.</b> Left & right rotation (LR, RR) absolute (unsigned) head repositioning accuracy (HRA) for each of the 10 repetitions at the retest session. ....	142
<b>Figure 39.</b> Left & right rotation (LR, RR) signed head repositioning accuracy (HRA) for each of the 10 repetitions at the retest session. ....	145
<b>Figure 40.</b> Bland and Altman Distribution plots for the LR $\Rightarrow$ 0 $\theta$ X-axis method agreement at session I. (A) all data. (B) effect with one outlier removed (outlier is circled in plot A) .....	148
<b>Figure 41.</b> A single HRA repetition position time plot for a single subject. (A) the Labview™ application measurement analysis window with cursors; position-time plot represented by yellow trace. (B) a plot of the same position-time plot. ....	151
<b>Figure 42.</b> Diagrammatic representation of study design.....	158
<b>Figure 43.</b> Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. ....	165
<b>Figure 44.</b> Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. ....	166
<b>Figure 45.</b> Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. ....	167
<b>Figure 46.</b> Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. ....	168
<b>Figure 47.</b> Plots of the absolute difference against mean (repeated measures) between Zebris and Laser methods by Examiner 1 for the head repositioning accuracy (HRA). ....	173
<b>Figure 48.</b> Plots of the absolute difference against mean (repeated measures) between Zebris and Laser methods by Examiner 2 for the head repositioning accuracy (HRA). ....	174
<b>Figure 49.</b> Plots of the difference against mean (repeated measures) between Zebris and Laser methods by Examiner 1 for the signed data head repositioning accuracy (HRA). ....	177
<b>Figure 50.</b> Plots of the difference against mean (repeated measures) between Zebris and Laser methods by Examiner 2 for the signed head repositioning accuracy (HRA) .....	178
<b>Figure 51.</b> Overview of study design and data handling for analysis of test-retest precision ...	180
<b>Figure 52.</b> Overview of study design and data handling for analysis.....	185
<b>Figure 53.</b> Superior view of horizontal ( $\theta$ Y) head repositioning using the laser pointer.....	195
<b>Figure 54.</b> Scatterplots showing the head repositioning accuracy (HRA) for the neck pain patients and control subjects.....	211
<b>Figure 55.</b> Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H).. ....	213
<b>Figure 56.</b> Scatterplots showing the head repositioning accuracy (HRA) with feedback for the neck pain patients and healthy subjects. All data are presented as degrees.....	215
<b>Figure 57.</b> Box plots of the signed repositioning error representing over/undershoot for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H).. ....	217



<b>Figure 58.</b> Scatterplots showing the head positioning accuracy to non-straight ahead targets (HPA-NSA) for the neck pain patients and healthy subjects..	219
<b>Figure 59.</b> Scatterplots showing the head positioning accuracy to non straight ahead targets (HPA-NSA) for the neck pain patients and healthy subjects.	220
<b>Figure 60.</b> Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H)..	222
<b>Figure 61.</b> Bland and Altman distribution plots with data exhibiting a relationship between the mean and difference: A – proportional bias; B – heteroscedasticity	256
<b>Figure 62.</b> A – Regression of difference of methods on the mean of two methods. B – Scatterplot to graphically define residuals; this is reflected by the vertical lines from the line of best fit to the observed measurements	259
<b>Figure 63.</b> A – Modified limits of agreement with equations for heteroscedastic data with no proportional bias ( $D = d$ ). B – Modified limits of agreement with equations for data exhibiting proportional bias but no heteroscedasticity.	261
<b>Figure 64.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the LR $\Rightarrow$ 0 trial.	265
<b>Figure 65.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the LR $\Rightarrow$ 0 trial.	266
<b>Figure 66.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the RR $\Rightarrow$ 0 trial.	267
<b>Figure 67.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the RR $\Rightarrow$ 0 trial.	268
<b>Figure 68.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Ext $\Rightarrow$ 0 trial.	269
<b>Figure 69.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the Ext $\Rightarrow$ 0 trial.	270
<b>Figure 70.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Flex $\Rightarrow$ 0 trial.	271
<b>Figure 71.</b> Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Flex $\Rightarrow$ 0 trial.	272
<b>Figure 72.</b> Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the LR $\Rightarrow$ 0 trial.	273
<b>Figure 73.</b> Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the RR $\Rightarrow$ 0 trial.	274
<b>Figure 74.</b> Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the Ext $\Rightarrow$ 0 trial.	275
<b>Figure 75.</b> Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the Flex $\Rightarrow$ 0 trial.	276

## List of Tables

<b>Table 1.</b> Investigations of cervicocephalic kinaesthetic sensibility utilizing repositioning accuracy to a straight ahead or neutral head position.....	17
<b>Table 2.</b> Investigations of cervicocephalic kinaesthetic sensibility utilizing repositioning and positioning tasks to non-neutral set points with the cervical range of motion.....	28
<b>Table 3.</b> Investigations of cervicocephalic kinaesthetic sensibility utilizing dynamic head movement control tasks. ....	32
<b>Table 4.</b> Inclusion & Exclusion criteria.....	36
<b>Table 5.</b> Comparison of active cervical range of motion (AROM) measurements (degrees) in neck pain (n=11) and control (n=11) subjects.....	44
<b>Table 6.</b> Comparison of absolute (unsigned) Global (R), Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) in neck pain (n=11) and control (n=11) subjects.....	49
<b>Table 7.</b> Comparison of signed Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) neck pain (n=11) and control (n=11) subjects <sup>†</sup> .....	50
<b>Table 8.</b> Guidelines for developing, describing and promoting the adoption of a new clinical test. ....	65
<b>Table 9.</b> Functions of clinical tests.....	66
<b>Table 10.</b> Terms used to describe common <i>reliability</i> study designs.....	68
<b>Table 11.</b> Sensitivity of statistical parameters to different types of error and bias. ....	76
<b>Table 12.</b> Effects of various biases and errors on statistical results from Pearson’s product moment correlation and Intra-class correlation coefficient analysis. ....	79
<b>Table 13.</b> A convenient data matrix, notational system and source of variance (factors) for data used in calculating Intraclass correlation coefficients. ....	87
<b>Table 14.</b> Effects of various biases and errors on statistical results from Ordinary Least Squares (OLS) Simple Linear Regression analysis. ....	101
<b>Table 15.</b> Effects of various biases and errors on statistical results from Bland and Altman method of analysis. ....	114
<b>Table 16.</b> Inclusion & Exclusion criteria.....	123
<b>Table 17.</b> Absolute (unsigned) vertical ( $\theta X$ ), horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =13) for trials in the horizontal movement plane (L & RR) conducted with the 5 and 10 repetition testing procedure. ....	135
<b>Table 18.</b> Signed horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =13) for trials in the horizontal movement plane (L & RR) conducted with the 5 and 10 repetition testing procedure.....	135
<b>Table 19.</b> Results of Bland and Altman analysis for agreement (n=13) <u>between</u> the 5 reps and 10 reps measurement methods at session I for the absolute X & Y-axis head repositioning accuracy (degrees). ....	137

<b>Table 20.</b> Results of Bland and Altman analysis for agreement (n=13) <u>between</u> the 5 reps and 10 reps measurement methods at session II for the absolute X & Y-axis head repositioning accuracy (degrees). .....	138
<b>Table 21.</b> Results of Bland and Altman analysis for agreement using repeated measurements <u>between</u> the 5 reps and 10 reps measurement methods for the absolute X & Y-axis head repositioning accuracy (degrees). .....	138
<b>Table 22.</b> Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=13) <u>between</u> the two different measurement methods (5 reps and 10 reps) with data sets exhibiting a relationship between the mean errors and differences. ....	138
<b>Table 23.</b> Results of Bland and Altman analysis for method agreement between two different measurement methods (5 reps and 10 reps <u>within trial</u> ) for the absolute X & Y-axis head repositioning accuracy (degrees). .....	142
<b>Table 24.</b> Results of Bland and Altman analysis for method agreement using between two different measurement methods (5 reps and 10 reps) at session I (test) for the signed Y-axis head repositioning accuracy (degrees). .....	143
<b>Table 25.</b> Results of Bland and Altman analysis for method agreement using between two different measurement methods (5 reps and 10 reps) at session II (retest) for the signed Y-axis head repositioning accuracy (degrees). .....	143
<b>Table 26.</b> Results of Bland and Altman analysis for method agreement using repeated measurements between two different measurement methods (5 reps and 10 reps) for the signed Y-axis head repositioning accuracy (degrees). .....	144
<b>Table 27.</b> Results of Bland and Altman analysis for method agreement between two different measurement methods (5 reps and 10 reps <u>within trial</u> ) for the signed Y-axis head repositioning accuracy (degrees). .....	144
<b>Table 28.</b> 'Reliability' findings for cervicocephalic kinaesthetic sensibility. ....	155
<b>Table 29.</b> Inclusion & Exclusion criteria .....	158
<b>Table 30.</b> Absolute (unsigned) vertical ( $\theta X$ ), horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =24) measured concurrently using the laser pointer and Zebris CMS70P method in the horizontal and vertical movement plane by two examiners at sessions two weeks apart. ....	169
<b>Table 31.</b> Signed repositioning error (degrees) in healthy subjects (n =24) measured concurrently using the laser pointer and Zebris CMS70P method in the horizontal and vertical movement plane by two examiners at sessions two weeks apart. These data represent the overshoot/undershoot characteristics in the plane of movement for HRA-SSA. ....	170
<b>Table 32.</b> Results of Bland and Altman analysis for method agreement using repeated measurements by examiner 1 between two different measurement methods (Zebris and Laser) for the absolute X & Y-axis head repositioning accuracy (degrees). .....	172
<b>Table 33.</b> Results of Bland and Altman analysis for method agreement using repeated measurements by Examiner 2 between two different measurement methods (Zebris and Laser) for the absolute X & Y-axis head repositioning accuracy (degrees). .....	172
<b>Table 34.</b> Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=24) between the two different measurement methods (Zebris and Laser) with data sets exhibiting a relationship between the mean errors and differences. ....	175

<b>Table 35.</b> Results of Bland and Altman analysis for method agreement using repeated measurements by Examiner 1 between two different measurement methods (Zebris and Laser) for the signed head repositioning accuracy (degrees) in the primary motion direction .....	176
<b>Table 36.</b> Results of Bland and Altman analysis for method agreement using repeated measurements by examiner 1 between two different measurement methods (Zebris and Laser) for the signed head repositioning accuracy (degrees) in the primary motion direction .....	176
<b>Table 37.</b> Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=24) between the two different measurement methods (Zebris and Laser) with data sets exhibiting a relationship between the mean errors and differences. ....	179
<b>Table 38.</b> Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 1 using the laser pointer method. ....	181
<b>Table 39.</b> Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 2 using the laser pointer method. ....	181
<b>Table 40.</b> Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 1 using the Zebris system.....	182
<b>Table 41.</b> Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 2 using the Zebris system.....	182
<b>Table 42.</b> Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 1 using the laser pointer method. ....	183
<b>Table 43.</b> Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 2 using the laser pointer method. ....	183
<b>Table 44.</b> Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 1 using the Zebris system.....	184
<b>Table 45.</b> Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 2 using the Zebris system.....	184
<b>Table 46.</b> Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session I using the laser pointer method. ....	186
<b>Table 47.</b> Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session II using the laser pointer method. ....	187
<b>Table 48.</b> Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session I using the Zebris system. ....	187
<b>Table 49.</b> Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session II using the Zebris system. ....	188
<b>Table 50.</b> Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 1 using the laser pointer method.....	189
<b>Table 51.</b> Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 2 using the laser pointer method.....	189
<b>Table 52.</b> Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 1 using the Zebris system. ....	189

<b>Table 53.</b> Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 2 using the Zebris system. ....	190
<b>Table 54.</b> General inclusion & exclusion criteria for both neck pain patient groups .....	205
<b>Table 55.</b> Group demographics and clinical characteristics. ....	209
<b>Table 56.</b> Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-SSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17). ....	210
<b>Table 57.</b> Comparison of signed repositioning error (degrees) for HRA-SSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	212
<b>Table 58.</b> Comparison of movement characteristics during HRA-SSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	214
<b>Table 59.</b> Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-SSA with feedback, in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17). ....	216
<b>Table 60.</b> Comparison of signed repositioning error (degrees) for HRA-SSA with feedback in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	216
<b>Table 61.</b> Comparison of movement characteristics during HRA-SSA with feedback in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	217
<b>Table 62.</b> Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17). ....	218
<b>Table 63.</b> Comparison of signed repositioning error (degrees) for HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	221
<b>Table 64.</b> Comparison of movement characteristics during HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction. ....	223
<b>Table 65.</b> A comparison of the present vs other studies HRA-SSA/NHP for healthy control subjects. Data represents absolute $\theta Y$ error for horizontal rotation repositioning tasks (Mean $\pm$ SD) .....	232
<b>Table 66.</b> A comparison of the present vs other studies HRA to non-straight ahead positions for healthy control subjects. Data represents absolute $\theta Y$ error for horizontal rotation repositioning tasks to 30° (Mean $\pm$ SD). ....	232
<b>Table 67.</b> Raw artificial data for agreement statistics examples .....	254

# Declaration of Authorship

I, .....

declare that the thesis entitled

.....  
.....

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

Rix, G.D.W. and Bagust, J. (2001). Cervicocephalic kinesthetic sensibility in patients with chronic, nontraumatic cervical spine pain. Arch Phys Med Rehabil 82 (7), 911-919.

Signed:.....

Date:.....

## Acknowledgements

There are many people I would very much like to acknowledge for their help, guidance and support. In many instances, without this support, the thesis would not have been completed.

I would like to express my sincere gratitude to my supervisor Professor Jeff Bagust for his patience, advice, feedback and invaluable support throughout course of study. In particular, I cannot thank him enough for his guidance and feedback during the final stages of this thesis – I could not have completed the work without this. He has shown me what scientific investigation should really be all about. I must also thank him for constructing the LabView analysis package enabling me to crunch the Zebris data with so much ease.

Special thanks also to my other supervisor, Professor Robert Walker for his support, interest and encouragement.

Many thanks also to the AECC research department for the funding and support.

Thank you to Dr. Alex Webb for your drive and support throughout this study but in particular your invaluable collaboration on our journey through the reliability statistics.

I greatly appreciate the help of Dr. Ian Brown for his help in developing the audio equipment and his invaluable advice regarding the experimental protocols.

A very grateful thanks must also go to Dr. Neil Osborne for being the second examiner in one of my studies and helping with the analysis of the data he collected.

I owe a special thanks to Mrs Angela Wood who allowed me to use the radiography room for my clinical data collection and was so helpful in accommodating me particularly when things were busy.

I must acknowledge and thank Dr. Kim Humphreys and Dr. Haymo Thiel for their invaluable help in screening the new patients for selection for my initial clinical study.

I must also offer a very special thanks to all my friends and colleagues at the AECC for their immense help and support when I was unwell and getting back to work. In particular, I must thank Dr. Edward Rothman, Mrs Rosemary Ashby, Mrs Lou Andrews, Dr. Ken Vall, Dr. Haymo Thiel, Dr. Jeff Bagust and Dr. Arvid Thorkeldsen.

Many thanks to the maintenance department at the AECC for building the laser pointer target.

Many thanks also to the University of Southampton for the support through the difficult times, in particular Mrs Lorraine Prout.

Lastly but certainly not least, I must thank my parents so much for the endless support, sacrifices and encouragement. This thesis would not have been possible without your care and support through my illness and recovery.

## Abbreviations

3-D	Three Dimensional
AE	Absolute Error
AECC	Anglo-European College of Chiropractic
BP	Blood Pressure
BSI	British Standards Institute
CI	Confidence Interval
CNS	Central Nervous System
CROM	Cervical Range of Motion Device
C-Spine	Cervical Spine
DNF	Deep Neck Flexion/Flexors
Ext	Extension
Ext $\Rightarrow$ 0	Extension Repositioning to Reference Zero
Flex	Flexion
Flex $\Rightarrow$ 0	Flexion Repositioning to Reference Zero
H <sub>0</sub>	Null Hypothesis
H <sub>1</sub>	Hypothesis
HRA	Head Repositioning Accuracy
Hz	Hertz
ICC	Intraclass Correlation Coefficient
KS	Kinaesthetic Sensibility
K-S	Kolmogorov-Smirnov
LLF	Left Lateral Flexion
LLF $\Rightarrow$ 0	Left Lateral Flexion Repositioning to Reference Zero
LoA	Limits of Agreement
LR	Left Rotation
LR $\Rightarrow$ 0	Left Rotation Repositioning to Reference Zero
MND	Motor Neurone Disease
MS	Multiple Sclerosis
MSD	Musculoskeletal Disorders
n	Sample Size
NDI	Neck Disability Index
NHP	Neutral/Natural Head Posture
NP-I	Neck Pain - Insidious Onset
NP-T	Neck Pain - Traumatic Onset
NRS	Numerical Rating Scale
nSA	Non-Straight Ahead
OLP	Ordinary Least Products
OLS	Ordinary Least Squares



PC	Personal Computer
PhD	Doctor of Philosophy
RA	Repositioning Accuracy
RLF	Right Lateral Flexion
RLF $\Rightarrow$ 0	Right Lateral Flexion Repositioning to Reference Zero
ROC	Receiver Operating Code
ROM	Range of motion
RR	Right Rotation
RR $\Rightarrow$ 0	Right Rotation Repositioning to Reference Zero
SD	Standard Deviation
S <sub>d</sub>	Standard Deviation (Bland & Altman Analysis)
SE	Signed Error or Standard Error
SMT	Spinal Manipulative Therapy
SPNTT	Smooth Pursuit Neck Torsion Test
SSA	Subjective Straight Ahead
US	Ultrasound
VAS	Visual Analogue Scale
WAD	Whiplash Associated Disorders
WLP	Weighted Least Products
Yrs	Years

# Dedication

This Thesis is dedicated to:

## My Parents and Sister

*For their patience, sacrifices, support, love  
and encouragement in helping me reach this  
privileged position.*

William J. Rix

Yvonne D. Rix

Alison M. Marsh

## And my Grandparents

*For their love, support, sacrifices and  
wisdom and to my grandfather for providing  
me with the inspiration to seize the  
educational opportunities that he never had.*

George E. Stickland

Dorothy F.A. Stickland

## **Chapter 1**

---

# **GENERAL INTRODUCTION AND AIMS OF THE RESEARCH**

---

### **1.1 HISTORY, BACKGROUND & STATEMENT OF THE PROBLEM**

Over the years it has become evident that neck pain and neck related disorders (including headache and radiating pain into the arms and upper back) are much more common than previously thought (Lidgren, 2008). The majority of people can expect to experience some degree of neck pain in their lifetime. For many, this will amount to nothing more than a mild discomfort which does not require treatment and which has no major impact on work or other activities. However, some will go on to develop prolonged or repetitive episodes of neck pain or neck related symptoms (Haldeman, Carroll and Cassidy, 2008). Although the individual risk of developing persistent neck pain is low, the number of affected persons in the general population is of concern (Haldeman, Carroll and Cassidy, 2008). Pain arising from the cervical spine has become a major cause of disability around the world (Lidgren, 2008). In addition, the cost of treatment for musculoskeletal disorders (MSDs), including neck pain and related disorders, is growing rapidly and taking a greater share of healthcare resources in both industrialised and developing countries (Haldeman, Carroll and Cassidy, 2008).

Neck pain is also one of the commonest reasons for visiting an accident and emergency department (Stussman, 1996), an ambulatory (out-patient basis) medical provider (Schappert, 1996) and the second most common complaint reported by patients seeking chiropractic care (Shekelle and Brook, 1991; Pedersen, Noddeskou and Wejse, 1992; Hurwitz, Coulter, Adams, Genovese and Shekelle, 1998). One study reported that as many as 44% of patients with chronic neck pain visit their general practitioner on a yearly basis (Borghouts, et al., 1999). There are several prevalence studies on neck pain in the general population (Hogg-Johnson, et al., 2008). These vary in quality. The majority of scientifically acceptable studies of adult populations are from North

America, Scandinavia, and northern European countries. Many of these studies focus on lifetime, 12-month or the 1-month prevalence of neck pain. As the period of time increases, the prevalence of neck pain generally increases (Hogg-Johnson et al., 2008). Depending on the definition of neck pain used, estimates of the 1-month prevalence for adults in the general population ranges from 15.4% to 45.3% (Kim, et al., 2001; Hagen, Einarsen, Zwart, Svebak and Bovim, 2002; Haldeman, Carroll, Cassidy, Schubert and Nygren, 2008; Hogg-Johnson et al., 2008). Again, depending on the neck pain definition used, the 12-month prevalence of neck pain range from 12.1% to 71.5% (Westerling and Jonsson, 1980; Ektor-Andersen, Isacsson, Lindgren and Orbaek, 1999; Haldeman, Carroll, Cassidy et al., 2008; Hogg-Johnson et al., 2008). Most estimates of 12-month prevalence in the general adult population range between 30% and 50% (Hogg-Johnson et al., 2008). Neck pain with associated limitation of activities (disability) is less common with 12-month prevalence estimates of 1.7% to 11.5% (Palmer, Syddall, Cooper and Coggon, 2003; Chiu and Leung, 2006; Haldeman, Carroll, Cassidy et al., 2008; Hogg-Johnson et al., 2008). In addition, measures of pain not qualified by frequency, duration and/or accompanying interference with activities tend to be larger than estimates for pain that have been qualified in some way (Haldeman, Carroll, Cassidy et al., 2008; Hogg-Johnson et al., 2008).

Considering the huge impact of neck pain on individuals, health care systems and society, there has been a considerable lack of systematic knowledge in this clinical field. The level of interest has also been well below the level appropriate for a topic of such relevance (Balague, 2008). Although some review articles have been published over the years (Ferrari and Russell, 2003; Devereaux, 2004; Murphy, 2004), only two major reviews of neck pain as a topic have emerged. For one subgroup of neck pain patients, Spitzer et al., (1995) published their *Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management*. Although subsequent studies have suggested some changes to the original publication (Hartling, Brison, Ardern and Pickett, 2001; Sterling, 2004) the Quebec Task Force guidelines have largely remained a gold standard for the classification and prognosis and treatment of *Whiplash Associated Disorders* (WAD). On February 15<sup>th</sup> 2008, the journal *Spine* published the long awaited 21 chapter report of *The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and its Associated Disorders*. This multidisciplinary international Task Force represents a unique gathering of international expertise

covering all relevant aspects related to traumatic ('whiplash') and non-traumatic onset 'mechanical' neck pain and its associated disorders (Rydevik, 2008). The goals of the Task Force were (Lidgren, 2008):

- to complete a systematic search and critical review of the scientific literature on neck pain and its associated disorders, including the epidemiology, diagnosis, prognosis, economic costs, and treatment of neck pain and its associated disorders
- to complete original research on the risks of neck pain
- to examine cost-effectiveness and patient preferences for various treatment options
- to collate the evidence, using best evidence synthesis, inform clinical practice for the management of neck pain and its associated disorders
- to indicate areas where further research is required.

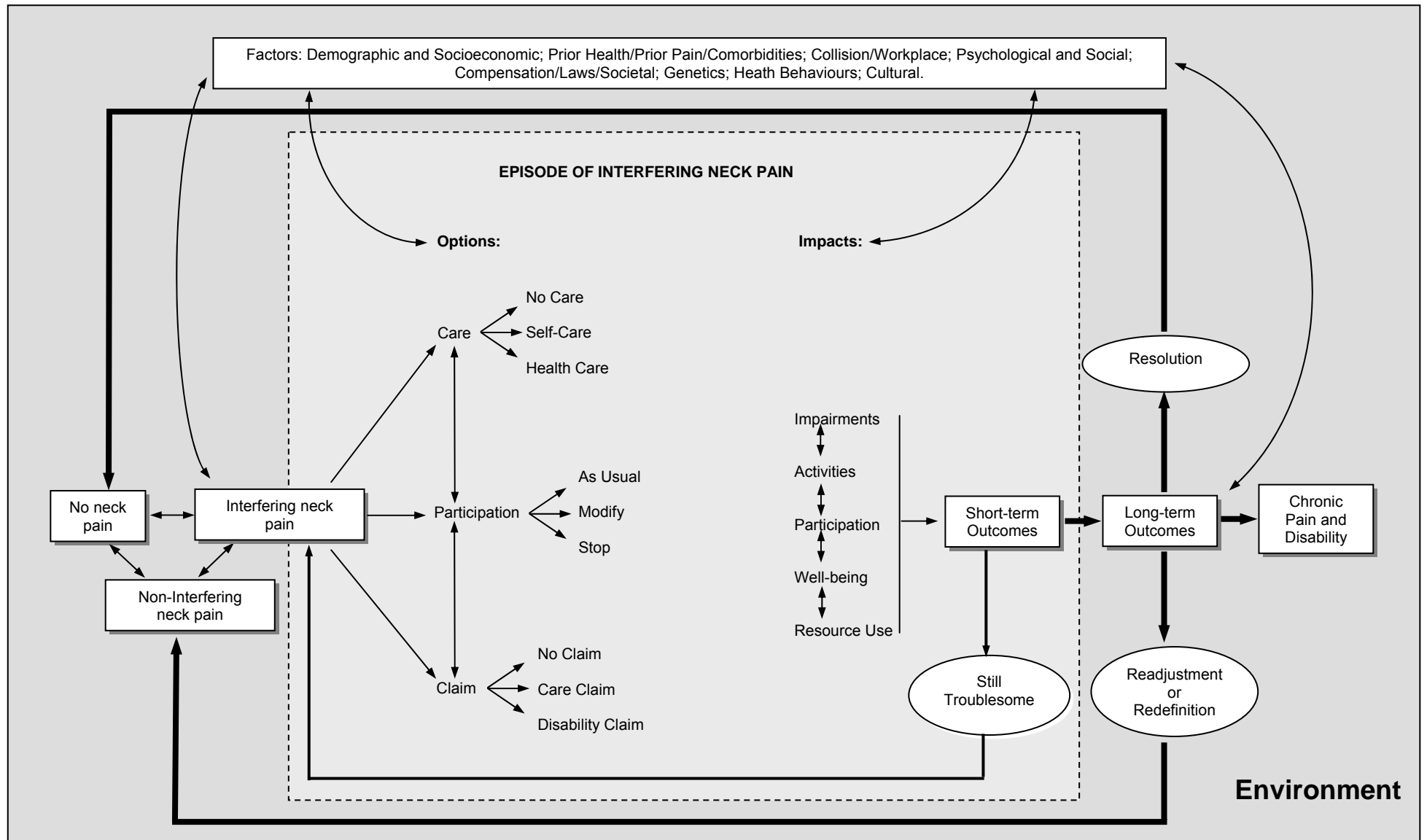
Along with these specific goals, this report should have a very significant impact on all health professionals who manage patients with neck pain and neck related disorders. It is hoped that it will change attitudes and beliefs about neck pain and its prevention, diagnosis, treatment and management (Haldeman, Carroll, Cassidy et al., 2008). It is also hoped that armed with this deeper understanding of neck pain and related disorders, clinicians will be better able to diagnose and treat people affected by these distressing and often disabling symptoms. It is also expected that the Neck Pain Task Force report will help highlight the need for more research and in particular the gaps in knowledge and areas of research that are more urgently needed (Lidgren, 2008).

Whether neck pain is likely to improve, reoccur, persist or worsen (the 'course' of neck pain or prognosis) is an extremely important question (Carroll, Hogg-Johnson, Van Der Velde, et al., 2008). Knowledge of the course of neck pain helps to guide expectations of people with neck pain and their health care providers. In addition, by knowing the usual course of recovery of neck related complaints, we can determine the effectiveness of interventions by whether they improve or worsen this natural course (Carroll, Hogg-

Johnson, Van Der Velde et al., 2008). Longitudinal prognostic studies in both adults and children, indicate that most people in the general population with neck pain do not experience a complete resolution of symptoms (Carroll, Hogg-Johnson, Van Der Velde et al., 2008; Haldeman, Carroll, Cassidy et al., 2008). Between 50% and 85% of those who experience neck pain at some initial point, will report neck pain again 1 to 5 years later (Mikkelsen, Salminen and Kautiainen, 1997; Mikkelsen, Sourander, Salminen, Kautiainen and Piha, 1999; Cote, Cassidy, Carroll and Kristman, 2004; Hill, Lewis, Papageorgiou, Dziedzic and Croft, 2004; Bot, et al., 2005; Pernold, Mortimer, Wiktorin, Tornqvist and Vingard, 2005; Carroll, Hogg-Johnson, Van Der Velde et al., 2008; Haldeman, Carroll, Cassidy et al., 2008). It appears that estimates in specific subpopulations of people with neck pain (workers and after road traffic accidents) are similar in size (Carroll, Hogg-Johnson, Cote, et al., 2008; Carroll, Holm, et al., 2008; Haldeman, Carroll, Cassidy et al., 2008).

One very useful outcome of *The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and its Associated Disorders* is the provision of a new conceptual model centred on the person with neck pain or who is at risk of neck pain (Guzman, et al., 2008). The main objective of this was to provide an integrated model for linking the epidemiology of 'mechanical' neck pain with its management and consequences (Figure 1) It also aimed to help organize and interpret existing knowledge and highlight gaps in the current literature (Guzman et al., 2008). The model essentially represents an overarching conceptual model and meaningful subgroups of people with neck pain (case definitions). One such meaningful categorisation is to group by onset e.g., 'whiplash' associated disorders (WAD) and insidious onset (non-traumatic) neck pain. This reflects different aetiological factors, and possible differences in associated complaints, pathology, perpetuating factors etc.

Although chronic neck pain can readily be defined in clinical terms (Figure 1), the underlying pathology remains largely unclear. Opinions vary widely on what causes neck pain and how to manage it (Haldeman, Carroll and Cassidy, 2008). It is felt that opinions on what causes or exacerbates neck pain depends often on the training and experience of the clinician than on scientific studies or consensus (Haldeman, Carroll and Cassidy, 2008).



**Figure 1.** Graphical representation of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and its Associated Disorders Conceptual Model for the onset, course, and care of neck pain. From Guzman et al., (2008).

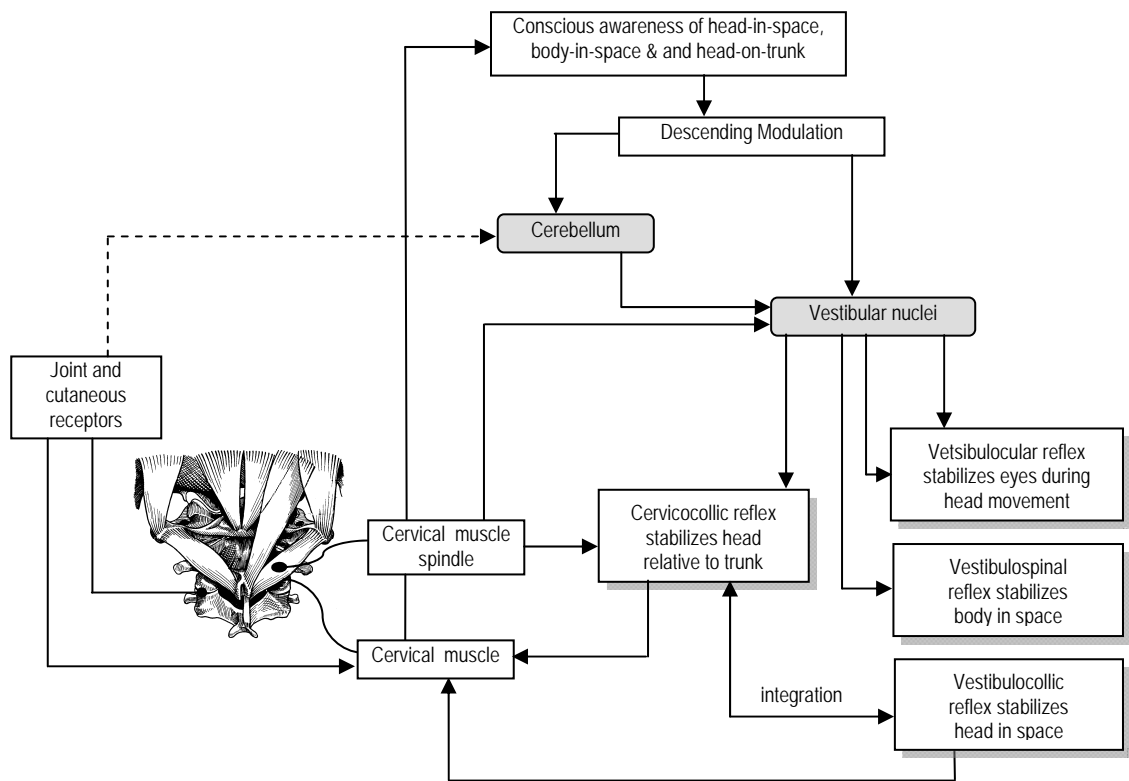
Cervical spine research has failed to demonstrate a consistent relationship between the presence of neck pain and pathoanatomy (Friedenberg and Miller, 1963; Heller, Stanley, Lewis-Jones and Heller, 1983; Gore, Sepic and Gardner, 1986; Boden, et al., 1990; Pettersson, Hildingsson, Toolanen, Fagerlund and Bjornebrink, 1994; Marchiori and Henderson, 1996; Ronnen, et al., 1996; Karlsborg, et al., 1997). There is no evidence that common degenerative changes in the cervical spine are a risk factor for neck pain (Peterson, Bolton, Wood and Humphreys, 2003; Haldeman, Carroll, Cassidy et al., 2008; Hogg-Johnson et al., 2008). As such, there has been an increasing interest in altered neuro-muscular-articular function in the pathogenesis of neck pain and other cervical spine related syndromes (Lewit, 1994; Murphy, 2000; Murphy, 2004; Liebenson, 2007).

## **1.2 FUNCTIONAL DEFICITS – CERVICAL SPINE PROPRIOCEPTION**

Over the last 17 years or so, the potential role of cervical proprioceptive (mechanoreceptive) dysfunction in the perpetuation of chronic neck pain has received increasing attention from researchers and clinicians. At the most simplistic level, neck mechanoreceptors are essential for sensing cervical spine movements and positions (Dutia, 1991; Bolton, 1998). According to the concept of a stabilizing system proposed by Panjabi (1992a; 1992b), integrity of the neuromuscular complex depends on three subsystems: passive structures, active structures and neuromuscular control. The last of these depends on adequate proprioceptive information. Clinically, the consequences of this altered proprioceptive sensibility or mechanoreceptive afferent integration and ‘tuning’ in the CNS, may be dysfunctional neuromuscular protection/stabilisation of articular tissues and control of head movement (Proske, Schaible and Schmidt, 1988). If the alteration in mechanoreception involves diminished sensibility, the result may also be a reduced modulation of cervical spine nociceptive activity (Wall, 1989; Seaman and Winterstein, 1998). Cervical mechanoreceptive input is also part of a large and complex array of integrated afferent information being presented to the central nervous system (CNS) from the entire locomotor system as well as other afferent systems such as sight, hearing, vestibular stimulation (Dutia, 1991). As such, neck mechanoreceptive input and CNS integration may also have a functional significance in coordination of eye, head and body movements as well as spatial orientation, self-motion perception and balance



(Figure 2). These relatively poorly understood functional connections might be one explanation for the additional symptoms seen in patients experiencing neck pain such as dizziness, dysequilibrium and disturbances of vision (Brown, 1992; Brandt, 1996; Wrisley, Sparto, Whitney and Furman, 2000; Brandt and Bronstein, 2001).



**Figure 2.** Simplified schematic of the neural subsystems, reflexes and connections for head and neck position sense and motor control (Armstrong, McNair and Taylor, 2008).

Proprioception testing is rapidly emerging as a clinical tool for evaluating the degree and nature of mechanoreceptive impairment in disorders of the locomotor system and assessment of treatment outcome. Proprioceptive therapies are also increasingly being incorporated into rehabilitation programmes particularly for the extremities (Kibler and Livingston, 2001; Ergen and Ulkar, 2008). As such, testing methods which can discriminate normal from abnormal function and provide an accurate measure of change over time are of fundamental importance. The first step in the development of this line of investigation requires a method of proprioception measurement that is *reliable* and

*valid* and should ideally have apparatus and methodology that can be applied easily in a clinical setting (time-efficient and cost within reasonable limits) (Maffey-Ward, Jull and Wellington, 1996; Kristjansson, Dall'alba and Jull, 2001). Before discussing this further, a focused review of some key aspects of terminology and methods of proprioception measurement is warranted.

### **Proprioception and Kinaesthesia**

Proprioception and kinaesthesia are terms which are often used in this area of study (Lam, Jull and Treleaven, 1999; Koumantakis, Winstanley and Oldham, 2002). The term *kinaesthesia* was coined by Bastian in 1888 to describe the perception of sensations about the static position or velocity of movement of those parts of the body moved by skeletal muscles (whether voluntarily generated or imposed – active or passive). It was also used to describe perceived sensations about the forces generated during muscular contractions even when there is no movement of the body part (isometric contraction) (McCloskey, 1978). More recently, kinaesthesia has been defined as a conscious sensation comprising three main components; the sensations of position and movement of joints; the sensation of force, effort and heaviness associated with muscular contractions; and the sensations of perceived timings of muscular contractions (Gandevia, McCloskey and Burke, 1992). The term *proprioception* was originally coined by Sherrington in 1906 to describe more broadly the conscious and unconscious afferent signals that originate in receptors that are stimulated by an organism's own movement. In this wider sense, the term encompasses vestibular and visual inputs as well as afferent input from muscles and joints. From these definitions, kinaesthesia could be seen as a submodality of proprioception.

More current descriptions and uses of the term proprioception refer only to the conscious sense of body position and movement arising from receptors in somatic tissues namely skin, muscles, tendons and joints. In this narrower sense, proprioception is widely described as a complex somatic entity encompassing several different components such as sense of position, velocity, movement detection and force (Lonn, Crenshaw, Djupsjobacka, Pedersen and Johansson, 2000; Swinkels and Dolan, 2000). From this perspective, kinaesthesia and proprioception could be seen as one of the same.

The terms are now used interchangeably to encompass the somatic sense of position and movement of the body parts (Proske et al., 1988; Swinkels and Dolan, 2000).

### **Investigation of Kinaesthetic Sensibility**

For this discussion and the remainder of the thesis, the term kinaesthetic sensibility will be used to describe the consciously perceived static and dynamic mechanoreceptive sensation arising from somatic tissues. Although there is probably a seamless integration of all aspects of mechanoreceptivity with functional postures and movements, for the purposes of investigation, the static positional and dynamic components of kinaesthetic sensibility are most often studied in isolation. Kinaesthetic sensibility is classically measured directly using two types of tests designed to assess position or movement sense. Position kinaesthetic sensibility (position sense) is most commonly measured by using some type of position matching procedure. This may involve measurement of repositioning accuracy (RA) to a subjective reference position actively located by the subject or, a procedure in which a target position is presented or controlled by the examiner and the subject must match that position. Repositioning protocols have included (Lonn, Crenshaw, Djupsjobacka and Johansson, 2000).

- 1) Passive presentation/passive replication (target position and replication of position by passive movement of body part).
- 2) Passive presentation/active replication (target position by passive movement of body part – reposition position by active movement of body part).
- 3) Active presentation/active replication (target position and replication of position by active movement of body part).

Comparisons of position matching data are often made using the absolute error; the difference between the target angle and the subject's estimated angle without examining the direction of error (unsigned values). This is often presented as the separate error components in orthogonal axes, the error component in the plane of movement for the kinaesthetic task or a global error representing the absolute direct distance from the target (distance resolved from the orthogonal error coordinates). Position sense is also assessed using the signed error; the values representing the tendency of the subjects to undershoot or overshoot the target angle. Movement or dynamic components of

kinaesthetic sensibility are most often investigated by determining the threshold to detect movement and/or the motion direction where the challenge is applied as a constant velocity movement or as a constant stimulus. The majority of direct measurement somatic kinaesthetic sensibility data has been obtained from studies involving the extremities. The majority of these studies have looked at movement sensibility. However, a relatively small but increasing body of knowledge is emerging on axial (spinal) kinaesthetic performance in healthy subjects and patients with spinal pain and related disorders.

### **Investigation of Cervical Spine Kinaesthetic Sensibility**

The tests aimed at assessing cervical spine afferent function in previous studies can be categorised into three broad groups according to the task utilised; 1) simple target-matching (repositioning) to a subjective straight ahead (SSA) or neutral/natural head posture (NHP) (summarised in Table 1 - end of chapter), 2) repositioning to a non-neutral set point within the cervical range of motion (summarised in Table 2) or 3) dynamic tests of head movement accuracy and control (summarised in Table 3).

The most commonly utilized kinaesthetic task has been repositioning to a NHP (SSA). However, currently there is still no solidly established measurement method for the routine evaluation of mechanoreceptive somatosensory function related to mechanical disorders of the cervical spine. Nearly all the investigations attempting to quantify cervical mechanoreceptive function have utilised kinaesthetic performance tests involving movements of the head on a stationary trunk (head-on-trunk). Two studies have incorporated tests involving movement of the trunk with the head stationary (Kristjansson et al., 2001; Kristjansson, Dall'alba and Jull, 2003). From the outset, one potentially limiting/confounding aspect of head-on-trunk testing methods must be recognised. If it is assumed that conscious, purposeful head movement kinaesthetic tasks (with vision occluded) are reliant on peripheral afferent information, then head-on-trunk movement and positioning tests potentially utilize information from both cervical and vestibular stimulation. Although a number of experimental arguments point to the predominance of a cervical proprioceptive role, the performance of these tests may not represent a specific measure of cervical spine afferent function. This is particularly

pertinent with subjects who have experienced head trauma (direct or indirect) and in those experiencing dizziness and/or dysequilibrium.

The isolation of one subsystem or region of the neuraxis may be even more problematic when the integrated nature and functional overlap of the various sensory inputs within the systems involved in equilibrium, spatial and self-motion awareness is considered. The presence of chronic nociception and possible subsequent CNS neuroplastic changes in areas of the neuraxis involved in spatial awareness and movement control may add a further layer of complexity to the problem (Apkarian, et al., 2004). Partly for these reasons, the term cervicocephalic kinaesthetic sensibility has often been used in previous investigations measuring kinaesthetic performance using head-on-trunk movement and positioning tests.

Investigations of cervicocephalic kinaesthetic sensibility (KS) have mainly consisted of the following study types:

- 1) Cross-sectional observational designs usually comparing neck pain patient groups with asymptomatic controls (Tables 1-3).
- 2) Interventional designs (experimental or clinical) mainly looking at the effects of various active and passive therapies on KS (Tables 1 and 2).
- 3) Studies examining the *reliability* of various KS testing methods and (Table 28 - Chapter 5).

What follows is a broad overview of what the investigations have demonstrated. More specific aspects of these will be appraised, revisited and developed through the following chapters of the thesis.

With one exception (Armstrong, McNair and Williams, 2005), studies employing repositioning to NHP protocols to compare kinaesthetic performance between neck pain patients and 'healthy' or 'asymptomatic' control populations have revealed increased head repositioning errors with patients currently experiencing chronic neck pain (Table 1) (Revel, Andre-Deshays and Minguet, 1991; Revel, Minguet, Gergoy, Vaillant and Manuel, 1994; Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998; Humphreys and Irgens, 2002; Kristjansson et al., 2003; Sterling, Jull, Vicenzino, Kenardy and

Darnell, 2003; Treleaven, Jull and Sterling, 2003; Feipel, Salvia, Klein and Rooze, 2006; Lee, Wang, Yao and Wang, 2007; Sjolander, Michaelson, Jaric and Djupsjobacka, 2008; Treleaven, et al., 2008) and with suspected somatosensory dizziness (Heikkila, Johansson and Wenngren, 2000; Treleaven et al., 2003; Treleaven et al., 2008).

However, a comparison across these studies revealed that the results had been variable and sometimes inconsistent. In some studies, the tests have also not been very discriminative between patient and control populations possibly highlighting the heterogeneity of subgroups of neck pain patients. The variations between studies may reflect different clinical characteristics in the neck pain populations investigated (e.g., onset – ‘whiplash’ vs insidious; severity of ‘whiplash’ etc.).

Consideration of the different methods used for testing may also explain the discrepancies between studies. These methodological differences together with differences in error variables presented and sometimes incomplete description of the group characteristics and method of testing makes comparison of the kinaesthetic deficits and pooling of results somewhat difficult. This is particularly problematic when trying to determine what constitutes normal cervicocephalic kinaesthetic sensibility using the data from control groups. Four studies have compared patients with current chronic neck pain with healthy control subjects utilizing head repositioning tests to one or more non-neutral set points within a plane of head movement (Loudon, Ruhl and Field, 1997; Kristjansson et al., 2003; Armstrong et al., 2005; Feipel et al., 2006). These investigations have yielded conflicting results (Table 2). In three studies, the repositioning errors have been larger than for control subjects (Loudon et al., 1997; Kristjansson et al., 2003; Feipel et al., 2006). However the differences have been small in two studies raising the question of clinical significance (Kristjansson et al., 2003; Feipel et al., 2006). Although similar tasks were performed in three of these studies (Loudon et al., 1997; Kristjansson et al., 2003; Feipel et al., 2006) variations in testing protocol and differences in measurement equipment and group characteristics (patient heterogeneity) are again the most obvious reasons for the contrasting data.

Studies examining cervicocephalic kinaesthetic sensibility using dynamic head movement control tasks have also produced conflicting results when comparing the performance of patients and healthy control subjects (Table 3). No difference was seen between patient subgroups (insidious onset neck pain and ‘whiplash’) and the healthy

control subjects when measuring the accuracy subjects could pass through a NHP whilst performing a figure-8 movement (Kristjansson et al., 2003). Using a computer generated random path of movement of the head ('the fly') as a target for subjects to track with head movements, the same investigators did find greater tracking errors for another group of 'whiplash' patients compared with healthy controls (Kristjansson, Hardardottir, Asmundardottir and Gudmundsson, 2004). As with the previous studies, differences in 'whiplash' patient characteristics (gender and current pain severity in particular) may also have been a contributing variable to the differences seen with these two testing approaches. The tracking test was also conducted with the eyes open raising the possibility that this test may also have examined aspects of altered eye movement control as opposed to cervical spine positional error. As highlighted earlier, altered eye movement control has been demonstrated in patients with persistent pain following 'whiplash' injury (Kelders, et al., 2005; Montfoort, et al., 2006).

### **Effects of Clinical Interventions on Cervicocephalic Sensibility**

Interventional (clinical) studies suggest positive effects on head repositioning accuracy to NHP in chronic neck pain patients ('whiplash' and mixed onset) using spinal manipulative therapy (SMT) (Rogers, 1997), active rehabilitation exercises (Heikkila and Astrom, 1996; Soderlund, Olerud and Lindberg, 2000; Humphreys and Irgens, 2002; Armstrong et al., 2005) and SMT combined with other passive therapies and active exercises (Palmgren, Sandstrom, Lundqvist and Heikkila, 2006). However with acute neck pain subjects, no significant changes have been demonstrated with "specific" neck mobilisation treatment (McNair, Portero, Chiquet, Mawston and Lavaste, 2006) and rehabilitation exercise (Soderlund et al., 2000). In both these studies, a case report (McNair et al., 2006) and a randomised controlled trial (Soderlund et al., 2000), it is difficult to know if the patients had impaired KS before treatment, as no asymptomatic/healthy control subjects were tested. In addition, as the methods of measurement were different to those used in other studies, no valid comparisons of error size between subject groups can be made. This further highlights the difficulty mentioned earlier with trying to contrast and pool data when such variations exist in the testing methods and equipment even when the same broad functional task is under investigation (e.g., repositioning to NHP).

The results from these interventional studies also serve to raise the issue of the functional relationship of cervicocephalic KS to other cervical spine clinical variables such as perception of pain, range of motion (ROM) etc. Improvements in KS with treatment directly applied to the neck (e.g., SMT) support the *validity* of the testing in that the tests may be measuring an aspect of kinaesthetic awareness related to the neck complaints. If it is assumed that the tests are a measure of cervical spine proprioceptive function, it is uncertain if the KS impairments seen are a primary neurophysiological component (functional impairment of receptors or altered CNS afferent tuning/integration) requiring independent/direct treatment approaches or whether the impairments are secondary to nociception, the perception of pain, altered joint mechanics, myopathology or other functionally interrelated components of neck dysfunction; in essence it is merely a functional symptom of the complaint not a causative factor. The results from the treatment of acute pain patients would suggest that deficits are probably not directly related to acute nociception and/or any secondary functionally related phenomenon such as local muscle spasm and joint restriction.

The relationship to chronic pain could be different. The persistence of pain may lead to more functional impairment of the neck of a different nature to that seen with more acute problems. It is also well known that functional neurophysiological changes such as central sensitization can be seen throughout the neuraxis with prolonged nociception (Vadivelu and Sinatra, 2005; Harvey and Dickenson, 2008). It is possible that these aspects of neuroplasticity cause secondary kinaesthetic deficits which may in themselves perpetuate a neck complaint but would be treated by treating the chronic pain directly. Most of the studies using active rehabilitation with chronic neck pain patients have utilised exercises that could be seen as primarily kinaesthetic in nature (Tables 1 & 2). Some have utilised specific eye-neck coordination, gaze stability and head positioning exercises (Revel et al., 1994; Humphreys and Irgens, 2002), and a ‘body awareness’ retraining (Heikkila and Astrom, 1996). All these active approaches have also simultaneously involved active head-on-trunk movements. It is not clear if the improvements seen resulted from a simultaneous general mobilization of the neck or if a more specific proprioceptive orientated training programme is needed. It is also interesting to note that in patients with persistent pain from a ‘whiplash’ injury, measurements of HRA-NHP were shown to have a relatively high predictive value (although low sensitivity) for abnormalities with the smooth pursuit neck torsion test



SPNTT and balance (postural control) (Treleaven, Jull and Lowchoy, 2006). It is therefore possible that the subjects in the previous rehabilitation exercise studies also had deficits in SPNTT and balance control. Treatment aimed at eye-head control may have simultaneously addressed the majority of functional problems (Treleaven et al., 2006). The possibly complex functional inter-relationship between cervicocephalic kinaesthetic sensibility parameters, other proprioceptive related variables and cervical pain and dysfunction will be explored in more detail in the general discussion chapter (Chapter 7).

### **‘Reliability’ of Cervicocephalic Tests**

Several studies have examined the *reliability* of various cervicocephalic KS testing methods (Table 28). They have all focused on the assessment of intra-examiner/test-retest *reliability* although three studies also incorporated an evaluation of inter-examiner *reliability* as well. With one exception (Kristjansson et al., 2004) the investigations were carried out solely in healthy, asymptomatic populations. Overall, it is difficult to draw solid conclusions on the *reliability* of any one method of testing partly due to the number of studies available but also due to the variation in tests examined, study designs and particularly the quality of methods/results reporting and appropriateness and quality of statistical analyses utilised. A more specific and detailed appraisal of the results and discussion of related issues of testing characteristics will follow in Chapters 3, 4 and 5 of the thesis.

## **1.3 AIMS AND OBJECTIVES OF THE RESEARCH**

Before commencing the course of study for this thesis, only seven studies had been published in the area of cervicocephalic kinaesthetic sensibility (Tables 1 and 2). As already highlighted, several of these found diminished cervicocephalic kinaesthesia in patients suffering chronic neck pain where the cause was not stated (Revel et al., 1991), mixed onset (Revel et al., 1994); or a cervical ‘whiplash’ injury was specifically involved (Heikkila and Astrom, 1996; Loudon et al., 1997; Heikkila and Wenngren, 1998). Evidence was also available to suggest that spinal manipulative therapy and active rehabilitation exercises may have a positive effect on KS (Revel et al., 1994; Heikkila and Astrom, 1996; Rogers, 1997).

Although a few studies had included ‘healthy’ or ‘asymptomatic’ subjects (Revel et al., 1991; Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998), before this PhD course of study, only one study had specifically examined KS performance in normal subjects (Christensen and Nilsson, 1999). Most studies had utilized a head-on-trunk repositioning task to SSA (NHP) using a laser pointer and target ahead of the subject but one had investigated performance with this task using more sophisticated 3-D motion analysis equipment (Christensen and Nilsson, 1999). The one study that examined the ability of patients to relocate the head to various rotation and side bending positions utilised the cervical range of motion device (CROM – see Table 1) (Loudon *et al.*, 1997). The clinical characteristics of the tests, in particular its ‘*reliability*’ had not been adequately established (Table 28).

At this point in the investigation of cervicocephalic kinaesthetic sensibility, several lines of investigation were identified as areas of interest for further research:

- 1) The specific comparison of KS sensibility in various patient subgroups (e.g., insidious vs ‘whiplash’ onset neck pain).
- 2) The development of more sophisticated methods of measuring head movement and repositioning errors which may also provide other possibly helpful kinematic data.
- 3) Further establishing the clinical characteristics of the tests in particular method agreement and ‘reliability’
- 4) Comparison of the relative discriminative value of repositioning tasks to both NHP and to non-neutral set points with the cervical range of motion.

The broad aims of this thesis were therefore to explore these focused areas of interest using appropriate investigations with a view to developing further our understanding and knowledge of cervicocephalic kinaesthetic sensibility.

**Table 1.** Investigations of cervicocephalic kinaesthetic sensibility utilizing repositioning accuracy to a straight ahead or neutral head position.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
<sup>†</sup> Revel, 1991	HRA-SSA	Observation of KS Difference; Patient vs Healthy Control group	Patients: n=30; (10 men and 20 women, mean age, 45 yrs [range, 23 to 73])  Controls: n=30; (10 men and 20 women, mean age, 44 yrs [range, 21 to 72])	'Chronic Cervical Pain' (mechanical)  Duration (mean); 89 months (range, 2 months to 34 years)  Pain severity (average) on day of testing (VAS); 39mm (range, 0 to 80 mm).  (no more details reported although data recorded)	<sup>a</sup> Laser pointer  (mounted on cycling helmet)  90 cm from target	Vision Occluded  SSA → Maximum rotation of head → SSA  10 reps ('trials') for each HRA task -Head repositioned by investigator to SSA after each 'trial')  Measured in degrees  HRA Tasks - LR, RR, 5 mins rest, Ext, Flex ⇒ 0; same order for each subject	<u>Global HRA-SSA</u> – significant diff ( $P < 0.01$ ) between patients and controls ( $P \text{ error} > C$ ) for LR, RR, Ext, Flex; Within groups, no diff in global error between plane (vertical vs horizontal) and direction of positioning (e.g. LR vs RR).  <u>Absolute HRA-SSA</u> - significant diff ( $P < 0.01$ ) between patients and controls ( $P \text{ error} > C$ ) for both $\theta X$ & $\theta Y$ components for LR, RR, Ext, Flex; error always larger in movement plane than orthogonal  <u>Signed HRA-SSA (Overshoot/undershoot)</u> – Overshoot tendency for patients with all tasks (LR, RR, Ext, Flex); For healthy subjects, no over/undershoot for horizontal tasks (LR & RR), overshoot for Flex, undershoot for Ext; These characteristics were statistically significant ( $P < 0.01$ ) for patients with LR, RR & Flex and with Ext & Flex for controls.  <u>Other</u> – 4.5° discriminant value (89% with ROC analysis); no effect of time on HRA error except Flex (at end of 30 min session); no correlation between HRA and pain characteristics (e.g. duration, severity etc).
<sup>†</sup> Revel, 1994	HRA-SSA	1) Quantification of KS for chronic neck pain patients  2) RCT  <i>Rehab Group (RG)</i> 8 week 'proprioceptive' (eye-neck coordination) rehab + symptomatic Rx  vs <i>Control group (CG)</i> Symptomatic Rx	Patients: n=60; (9 men and 51 women, mean age $\pm$ SD, 48 $\pm$ 14 yrs [range, 25 to 80])  RG: n=30; (8 men and 22 women, median age, 45 yrs [range, 25 to 74 yrs])  CG: 30; (1 male and 29 women, median age, 46.5 yrs [range, 25 to 84 yrs])	'Chronic Neck pain' (mechanical)  Duration (median); 36 months (range, 3 months to 18 years)  Pain severity (VAS); 48.2 $\pm$ 23.8 (mean $\pm$ SD)  OA; 37% exhibited 'Obvious' degenerative change or isolated disc space narrowing  'minor injury' at onset; 28%	<sup>a</sup> Laser pointer  (mounted on cycling helmet)  90 cm from target	As per Revel 1991 with following differences and uncertainties  (Repositioned after each trial?)  HRA Tasks - LR, RR  Same order of tasks for each subject?	<u>Demographics &amp; Clinical/Pathological Characteristics</u> – NS differences between groups except higher number of women in CG  <u>Global HRA-SSA</u>  1) Mean ( $\pm$ SD) Patient HRA was 7.7° ( $\pm$ 3.3); 82% sensitivity.  2) Significant increase ( $P < 0.001$ ) in mean HRA for RG after treatment (7.5° $\pm$ 3.7 vs 5.5° $\pm$ 2.6); No diff in CG.  <u>Other</u> – Significant diff ( $P < 0.01$ ) between RG & CG in Pain intensity and full active ROM (RR to LR) after trial; Good or very good self reported functional improvement 60% in RG vs 8% in CG.  No significant diff in daily intake in of NSAID's, analgesic drugs & full active ROM from Flex to Ext.
<sup>†</sup> Heikkila, 1996	HRA-SSA	1) Observation of KS Difference; Patient vs Healthy Control group  2) Effect of Rehab treatment (Dropsy's body awareness) on Patient KS - <i>Rehab Group (RG)</i>	Patients: n=14; (7 men and 7 women, mean age, 36 yrs [range, 23 to 47])  RG; n=8  Controls: n=34; (11 men and 21 women, mean age, 35 yrs [range, 26 to 53])	'Chronic Whiplash Syndrome'  Duration; 6 months to 10 years	<sup>a</sup> Laser pointer  (mounted on helmet)  90 cm from target	As per Revel 1991 with following differences and uncertainties  'target repositioned (light beam on zero of target) after each trial'  Measured in cm's  Order of tasks (same as Revel '91)?  Same order of tasks for each subject?	<u>Global HRA-SSA</u>  1) Significant reduced ( $P < 0.001$ ) for patients vs controls for LR, RR, Ext, Flex; Within groups, no significant diff between plane and direction of positioning for controls and patients although vertical plane error > horizontal error for patient group.  2) Significant increase ( $P < 0.05$ ) in mean HRA for RG after treatment in RR and Ext (trend for all to increase)  <u>Signed HRA-SSA (Overshoot/undershoot)</u> – Overshoot tendency for patients and for healthy subjects with RR, Ext and Flex tasks; Undershoot tendency for both groups with LR task; Significant diff ( $P < 0.05$ ) between patients and controls for LR and most obviously with Flex ⇒ 0  <u>Other</u> – No significant difference in pain severity for RG after treatment; no correlation between age and KS

Table 1. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
†Rogers, 1996	HRA-SSA	Cohort Study (matched) Effect of SMT vs stretching exercises on KS	SMT Patients: n=10; (5 men and 5 women, mean age ± SD, 48 ± 13 yrs [range, 28 to 69])  Stretching Patients: n=10; (6 men and 4 women, mean age ± SD, 49 ± 16 yrs [range, 21 to 66])	'Chronic Neck pain' (mechanical) – consistent daily neck pain > 4 months  Duration (mean ± SD ); 16 ± 14 yrs (SMT) & 17 ± 13 yrs (Stretching)  Pain severity (VAS [mean ± SD]); 36 ± 27.5 (SMT), 40 ± 17 (Stretching).  Past neck trauma; 7/10 (SMT & Stretching)  'Associated' dizziness; 3/10 (SMT), 5/10 (Stretching).	<sup>a</sup> Laser pointer  (mounted on baseball cap)  120 cm from target	Vision Occluded SSA → Maximum rotation of head → SSA  1 rep for each HRA task – error's added to achieve one Global HRA score (No new reference position between tasks)  Measured in cm's  HRA Tasks - RR, LR, Flex, Ext ⇒ 0; same order for each subject?	<u>Global HRA-SSA</u> – significant diff ( $d = -14.5$ cm) for SMT group ( $P < 0.001$ ), 9/10 improved. For Stretching group 5 improved, 5 worsened – overall diff of -4 cm.  <u>Other</u> – VAS; 8/10 SMT patients improved, 6/10 stretching patients improved but 4/10 worsened. No statistically significant diff pre → post in both groups and between groups at each point in study.
†Heikkila, 1998	HRA-SSA	1) Observation of KS Differences; Patient vs Healthy Control group  2) Observation of Oculomotor & Active cervical range of motion Difference; Patient vs Healthy Control group	Patients: n=27; (14 men and 13 women, mean age, 38.8 yrs [range, 18 to 66])  Controls: n=39; (15 men and 24 women, mean age, 35 yrs [range, 26 to 53])  (NB: For Oculomotor tests control group was different - n=25; (median age, 34 yrs [range, 25 to 40])	'Chronic Whiplash Syndrome' (WAD grades II & III)	<sup>a</sup> Laser pointer  (mounted on helmet)  90 cm from target	As per Heikkila 1996 with following difference  No rest stated between vertical and horizontal positioning tasks.	<u>Global HRA-SSA</u> – significant diff ( $P < 0.001$ ) between patients and controls ( $P$ error > C) for LR, RR, Ext, Flex; Within groups, no diff in global error between plane (vertical vs horizontal) and direction of positioning (e.g. LR vs RR).  <u>Signed HRA-SSA (Overshoot/undershoot)</u> – Overshoot tendency for patients with (Ext, Flex); For healthy subjects, no meaningful over/undershoot for any tasks although a slight tendency to undershoot in 3/4; Significant diff ( $P < 0.01$ ) between patients and controls for vertical movement tasks  <u>Other</u> – 6 cm discriminant zone for normal vs abnormal HRA; slightly lower HRA in patients with no symptoms 2 years after whiplash; Positive correlation for HRA with age for 'whiplash' patients; no correlation between age and HRA for control group.
†Christensen, 1999	HRA-SSA	Observational study  To determine the ability of healthy asymptomatic subjects to reproduce the neutral zero head position.	Subjects: n=38; (20 men and 18 women, mean age, 24.3 yrs [range, 20 to 30])	'Asymptomatic' Subjects	<sup>b</sup> CA 6000 Spine Motion Analyser	Eyes closed SSA → Movement of head in all 6 movement directions for 5 secs → SSA  3 repetitions  Measured in degrees	<u>Absolute HRA-SSA</u> (mean ± SD)  θX component (Flex/ext plane), $2.7^\circ \pm 2.1$ (range, 0 to 7.2)  θY component (Rotation plane), $1.0^\circ \pm 0.85$ (range, 0 to 3.3)  θZ component (Lateral flexion plane), $0.65^\circ \pm 0.67$ (range, 0 to 2.9)
Heikkila, 2000	HRA-SSA	Pilot Study - Single subject experimental design.  To compare effects of SMT, acupuncture, NSAID gel & no therapy on KS, dizziness, pain & ROM	Patients: n=14; (6 men and 8 women, mean age, 36 yrs [range, 22 to 54])  'Healthy Controls for KS test: n=39; (15 men and 24 women, mean age, 35 yrs [range, 26 to 53])	'Vertigo of suspected cervical origin'  All screened by ENT specialist and Physical medicine specialist (for presence of neck dysfunction)	<sup>a</sup> Laser pointer  (mounted on cycling helmet)  90 cm from target	As per Heikkila 1996	<u>Global HRA-SSA</u>  Significant diff ( $P < 0.001$ ) between patients and controls ( $P$ error > C) for LR, RR, Ext, Flex. Greatest difference for HRA tasks in the vertical plane.  Significant improvement ( $P < 0.05$ ) in vertical HRA after SMT and Acupuncture  Reduction in dizziness symptoms after SMT and Acupuncture (no diff between therapies); No pain relief with SMT but significant reduction with Acupuncture & NSAID gel.

**Table 1. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Soderlund, 2000	HRA-SSA  (really a visually reference to a target in front of subject)	RCT  Effect of Regular Exercise treatment (ROM & breathing shoulder relaxation exercise) on KS (& pain, ROM)  vs  Regular Exercise treatment and neck movement coordination exercise	<u>Patients</u> (total): n=59; (M:F, 24:35; mean age 34 yrs)	'Acute Whiplash Injury'  Arising from acceleration-deceleration movement of the head without direct head trauma (most related to RTA).  Pts referred from ER to Ortho Clinic  14% WAD I  83% WAD II  3% WAD III	<sup>a</sup> Laser pointer  (mounted on helmet)  50 cm from archery target	Eyes closed (for reposition task)  Eyes open → rotation of head → 'centre': Eyes closed → rotation of head → 'centre'  ? reps  Measured in mm's  HRA Tasks – RR & LR⇒ 0 (same order for each subject?)	<u>Global HRA-SSA</u>  HRA for both movement tasks generally improved slightly over 6 months for all groups. No statistically significant difference over time for either group.
Kristjansson, 2001	HRA-SSA  (also included other KS tasks – see next table)	Test-Retest 'Reliability'  (Study primarily aimed at assessing reliability of 5 different KS tests in healthy subjects – reliability results reported in Chapter 5)	Subjects: n=19; (7 men and 12 women, mean age ± 2SD, 35.1 ± 10 yrs).	'Healthy' Subjects  No current or prior history of musculoskeletal pain in neck or upper limbs.	<sup>c</sup> 3-Space Fastrak	Vision Occluded for tests  3 repetitions  Measured in degrees  Tasks; RR & LR for 1-3  <u>Test 1</u> - As per Revel 1991 with above rep diff <u>Test 2</u> - 30° L or RR of head (examiner positioned) → SSA <u>Test 3</u> - 30° L or RR of trunk (examiner positioned) → SSA <u>Test 4</u> - fig 8 movement of head x3 → SSA	<u>Absolute HRA-SSA</u> (mean ± SD)  θY component (axial/horizontal rotation plane) used for analysis  Test 1) 2.46° ± 1.32 Test 2) 3.46° ± 1.32 Test 3) 5.95° ± 4.10 Test 4) 2.43° ± 4.10  A statistically significant difference ( <i>P</i> < 0.001) was seen between all the different Tests excepts Test 1 vs Test 4
Humphreys, 2002	HRA-SSA  (Really assessed NHP)	1) Observation of KS  Difference; Patient vs Asymptomatic Control group  2) Stratified RCT  Effect of Rehab Exercises (eye-head-neck coordination) on KS  4 grps: Symptomatic + ex (SEG); Symptomatic no ex (SNEG). Asympt + ex (AEG); Asympt no ex (ANEG).	<u>Patients</u> (total): n=28; (M:F, 14:14; mean age 22.6 yrs [range, 19 to 30])  <u>Asymptomatics</u> (total): n=28; (M:F, 14:14; mean age 23.9 yrs [range, 19 to 31])	'Chronic Neck pain' (mechanical) – consistent daily neck pain > 3 months  2-3 episodes/wk  Duration (mean); 18 mths (range, 5 months to 12 years)  Past whiplash; 17/28 (61%); 10 in SEG & 7 in SNEG.  Intermittent dizziness: 10/28 (36%); 7 in SEG & 3 in SNEG.	<sup>a</sup> Laser pointer  (mounted on cycling helmet)  90 cm from target	As per Revel 1991 (see earlier) with following differences  Measured in cm's  HRA Tasks – Ext, LR, RR & Flex ⇒ 0  Not clear if same order used for each subject & at follow-up and whether new ref position found between tasks	<u>Global HRA-SSA</u>  1) Significant diff ( <i>P</i> < 0.001) between patients and controls ( <i>P</i> error > C) for LR, RR, Ext, Flex; Whiplash patients had significantly ( <i>P</i> < 0.05) diminished HRA particularly for LR & Flexion.  Within subgroups, no significant diff between plane and direction of positioning for controls and patients although vertical plane error > horizontal error for patient groups particularly SEG.  2) Significant improvement ( <i>P</i> < 0.001) in mean HRA for SEG in all test directions compared with other groups. Very little change observed for HRA in any of the other groups  <u>Other</u> – Significant reduction ( <i>P</i> < 0.001) in pain intensity for the SEG

Table 1. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Kristjansson, 2003	HRA-SSA  (Really assessed NHP)  (also included other KS tasks – see next table)	Observation of KS Difference; Patient subgroups (insidious onset & whiplash injury) vs Healthy Control group	Patients (insidious onset): n=20; (M:F, 11:9; mean age ± SD, 30.1 ± 9.1 yrs)  Patients (whiplash injury): n=22; (M:F, 11:11; mean age ± SD, 33.4 ± 10.6 yrs)  Asymptomatic: n=21; (M:F, 10:11; mean age ± SD, 26.9 ± 6.4 yrs)	'Chronic mechanical Cervical Pain': 3 & 48 mths duration  -Insidious Onset patients: Pain duration (mean ± SD); 28.6 ± 15.5 mths; current severity (VAS), 1.82 ± 2.0.  -Whiplash Injury patients: Pain duration (mean ± SD); 21.9 ± 12.5 mths; current severity (VAS), 3.37 ± 2.8.	°3-Space Fastrak	As per Kristjansson, 2001 (see above) with following changes:  Only used Tests 1 & 4 for NHP KS tasks  Random order of testing	<u>Absolute HRA-SSA</u> (mean ± SD)  θY component (axial/horizontal rotation plane) used for analysis  Although for both tests, the patient HRA was slightly worse than the asymptomatic groups, a statistically significant difference ( $P < 0.01$ ) was seen only for Test 1. Post-hoc analysis showed this difference to be between the asymptomatic group and both patient groups. No significant difference was seen between the insidious onset and whiplash patient groups.  NB: The whiplash Injury patients had significantly higher pain and disability scores than insidious group and exhibited slightly worse HRA in both tests.
Sterling, 2003	HRA-SSA  (Really assessed NHP)	A prospective longitudinal study  Observation of difference between Patient and control groups within 1 month of injury and 2 & 3 months post injury	Patients (Total): n=66; (M:F, 21:45; mean age ± SD, 36.27 ± 12.7 yrs)  Asymptomatic: n=20; (M:F, 8:12; mean age ± SD, 40.1 ± 13.6 yrs)	'Whiplash Injury'  WAD II & III following RTA (87% graded as WAD II)  NDI score used to classify patient subgroups at 3 months  - Recovered (<8 NDI)  - Mild Pain & Disability (10-28 NDI)  - Moderate/severe Pain & Disability (>30 NDI)  TAMPA score – fear of movement/re-injury.	°3-Space Fastrak	As per Revel 1991 (see earlier) with following changes  'comfortable' limits for movement in test.  Patient able to 're-align' NHP visually before each new task.  3 repetitions  HRA Tasks - LR, RR, & Ext ⇒ 0; same order for each subject  ROM & superficial neck flexor activity also measured	<u>Absolute HRA-SSA</u>  Error component in 1° movement planes used for analysis (e.g. θY component for L&RR)  At all time-points over the 3 months post injury, a significantly ( $P < 0.01$ ) worse RR KS was seen in the 'moderate/severe' patient group compared with other groups. Post-hoc analysis showed this difference to be between the asymptomatic and the 'moderate/severe' patient groups.  NB: There was no effect for age and gender on HRA; there was no change in HRA over time for any group.
Treleaven, 2003	HRA-SSA  (Really assessed NHP)	Observation of KS Differences; Patient subgroups (WAD no dizziness & WAD + dizziness) vs Asymptomatic Control group	Patients (WAD + dizzy): n=76; (71% female; mean age 39.1 yrs)  Patients (WAD no dizzy): n=26; (73% female; mean age 40.2 yrs)  Asymptomatics: n=44; (66% female; mean age 34.1 yrs)	'Chronic Whiplash Syndrome' (WAD grades II & III) > 3 mths (96 were WAD II)  -WAD D patients: Pain duration (mean); 1.6 yrs; pain at rest (VAS), 4.94; Neck Index Score, 55.3  - WAD ND patients: Pain duration (mean); 1.5 yrs; pain at rest (VAS), 3.96; Neck Index Score, 43.1	°3-Space Fastrak	As per Revel 1991 (see earlier) with following changes  One eyes open practice movement in each direction  3 repetitions  Patient able to visually 're-centre' NHP before each new task.  HRA Tasks - RR, LR, & Ext ⇒ 0; same order for each subject	<u>Absolute HRA-SSA</u> – significantly ( $P < 0.05$ ) greater errors in all testing directions for the overall WAD group vs controls in the 1° movement planes (e.g. θY component for L&RR); NB – WAD ND subjects had similar HRA (slightly larger error) vs controls for LR & RR ⇒ 0; WAD-D had worse HRA than WAD-ND for LR & RR ⇒ 0. Both patient groups contributed equally to diff vs controls with Ext ⇒ 0.  <u>Signed HRA-SSA (Overshoot/undershoot)</u> – Slight overshoot tendency for overall WAD group & controls with all tasks (LR, RR, Ext ⇒ 0); No differences between groups in L & RR rotation – WAD subjects more likely to overshoot with Ext ⇒ 0 (65% vs 42%)  <u>Other</u> – NB; not a clear separation between WAD and control groups.

Table 1. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Armstrong, 2005	HRA-SSA  (tested NHP)  (also included other KS tasks – see next table)	1) Observation of KS Difference; Patient vs Asymptomatic Control group  2) Effect of Deep neck Flexor(DNF) activation on KS  4 grps: Symptomatic + DNF(SDNF); symptomatic no DNF (SNDNF). Asympt + DNF(ADNF); Asympt no DNF (ANDNF).	<u>Patients</u> (total): n=23; (M:F, 8:15; mean age $\pm$ SD, 41.2 $\pm$ 11.9 yrs [range, 20.4 to 67.9])  <u>Asymptomatics</u> (total): n=23; (M:F, 10:13; mean age $\pm$ SD, 33.9 $\pm$ 12.1 yrs[range, 19.2 to 62.5])	'Whiplash Syndrome'  WAD II & III (87% graded as WAD II)  Duration of symptoms (mean $\pm$ SD); 31 $\pm$ 32 months.  NDI score (mean $\pm$ SD); 12 $\pm$ 5.5  PSFS score (mean $\pm$ SD); 4.7 $\pm$ 1.6	<sup>c</sup> 3-Space Fastrak*	Vision Occluded  NHP $\rightarrow$ Movement of head to self-selected mid position $\rightarrow$ NHP  3 repetitions  Measured in degrees (*at each receptor site)  HRA Tasks – Flex , Ext, LR and RR $\Rightarrow$ 0 (not clear on order within or between subjects)	<u>Absolute HRA-SSA</u> (mean of all specific movements 'pooled' together – plane of movement error component?) 1) No significant difference in HRA between whiplash & healthy subjects (data 'pooled' for intervention results) 2) No significant difference in HRA between DNF and no DNF groups <u>Signed HRA-SSA (Overshoot/undershoot)</u> 1) No significant difference in overshoot/undershoot characteristics between groups or tasks 2) No significant difference in overshoot/undershoot characteristics between DNF and no DNF groups NB: Irrespective of group, type of task or intervention, 65-80% of trials resulted in overshoot (all < 1.0°?) Other: No significant correlation between absolute HRA & NDI, PSFS, Active ROM; pain increased by 15% during the trials but no significant correlation between absolute HRA and pain severity and duration
Feipel, 2006	HRA-SSA  (tested NHP)  (also included other KS tasks – see next table)	Observation of KS Difference; Patient vs 'Healthy' Control group	Patients: n=29; (62% female; mean age $\pm$ SD, 37 $\pm$ 14 yrs)  Controls: n=26; (54% female; mean age $\pm$ SD, 35 $\pm$ 11 yrs)	'Whiplash Syndrome'  WAD I-III (traffic or sports injury); 29% graded as WAD I-II  Duration (mean time from accident $\pm$ SD); 31 $\pm$ 32 months.	<sup>b</sup> CA 6000 Spine Motion Analyser	Vision Occluded  NHP $\rightarrow$ Maximum Flexion to Extension head movement $\rightarrow$ NHP  4 reps  Measured in degrees	<u>Absolute HRA-SSA</u>  Significant ( $P < 0.05$ ) difference (WAD larger error) in HRA for WAD grp vs controls. However differences were small  For both groups, the repositioning error was largest in the plane of movement for the trial.
Owens, 2006	HRA-SSA  (Really assessed NHP)	Experimental study in asymptomatic subjects investigating the effects of contraction and/or shortening of neck muscle on KS.  1) "No hold" after active movements  2) "Passive hold"  3) "Active hold"	Subjects: n=48; (36 men and 12 women; mean age $\pm$ SD, 28.2 $\pm$ 4.8 yrs [range, 21 to 40])	'Asymptomatic' Subjects	<sup>b</sup> CA 6000 Spine Motion Analyser	Eyes closed for repositioning protocol  NHP (eyes open) $\rightarrow$ (eyes closed) 5 neck movements (Ext or LLF) $\rightarrow$ 1 of the 3 neck muscle conditioning sequences $\rightarrow$ NHP  All tests in random order	<u>Signed HRA-SSA (Overshoot/undershoot)</u>  Extension Conditioning: in plane of movement (sagittal, $\theta$ Y axis), undershoot tendency with conditions 1 & 2; overshoot tendency for condition 3 (Active hold) – this diff was statistically significant diff ( $P < 0.001$ ) compared with conditions 1 & 2. No under/shoot tendency in the other 2 orthogonal planes with all conditions.  LLF Conditioning: in plane of movement (frontal, $\theta$ Z axis), no under/shoot tendency with all conditions. In sagittal plane, overshoot tendency for condition 3 (Active hold) – this diff was statistically significant diff ( $P < 0.001$ ) compared with conditions 1 ("No hold").

**Table 1. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Palmgren, 2006	HRA-SSA  (Really assessed NHP)	RCT  Effect of Chiropractic Care (education, advice, SMT, soft tissue techniques & spine stabilizing exercises) on KS (& pain, ROM) vs No treatment other than education, advice and 'simple exercises'	Patients (treatment): n=18; (M:F, 7:11; mean age $\pm$ SD, 32.7 $\pm$ 8.2 yrs [range, 20 to 49])  Patients (control): n=19; (M:F, 6:13; mean age $\pm$ SD, 31.2 $\pm$ 9.0 yrs [range, 18 to 53])	'Chronic insidious onset Cervical Pain' (mechanical) > 12 weeks  Duration – not stated but 'continuous' pain  Pain severity – measured on VAS scale but descriptives for each group not reported	<sup>a</sup> Laser pointer  (mounted on ice hockey helmet)    100 cm from target	Vision Occluded  NHP determined by investigators  NHP → 'Sub-maximal' head rotation → NHP  Not clear on exact protocol of calibration, no of reps & order  Measured in cm's  Recorded instructions	<u>Global HRA-SSA</u>  Significant increase ( $P < 0.05$ ) in mean HRA for treatment group in all directions (RR, LR Ext, Flex, RLF, LLF)  No significant differences in control group except RR HRA  <u>Other</u> – No statistically significant differences in AROM in either group; 29mm drop in VAS for treatment group – no change for control group.
Treleaven, 2006	HRA-SSA  (Really assessed NHP)	Cross-sectional Observation & Correlation analysis  1) Association between tests of HRA, balance and smooth pursuit neck torsion (SPNT) in WAD patients  2) Usefulness of abnormal HRA in predicting abnormalities with the other tests.	Patients (WAD + dizzy): n=50; (76% female; mean age 35.5 yrs[range, 19 to 46])  Patients (WAD no dizzy): n=50; (76% female; mean age 35 yrs[range, 18 to 46])  'healthy control subjects': n=40; (58% female; mean age 29.6 yrs[range, 19 to 45]) Used to establish normative values for the secondary analysis.	'Persistent Whiplash Syndrome' > 3 mths  WAD II  -WAD D patients: duration since injury (mean); 1.4 yrs (range, 18 to 46); pain at rest (VAS), 4.94; Neck Index Score, 55.3  - WAD ND patients: Pain duration (mean); 1.6 yrs; pain at rest (VAS), 3.96; Neck Index Score, 43.1	<sup>c</sup> 3-Space Fastrak	As per Revel 1991 (see earlier) with following changes  One practice movement with eyes open  Patient able to 're-centre' NHP before each new task.  3 repetitions HRA Tasks - RR, LR, & Ext $\Rightarrow$ 0; same order for each subject?  Posturography & SPNT also measured  The order of each testing procedure was randomised for each subject.	<u>Absolute HRA-SSA</u>  Error component 1° movement planes used for analysis(e.g. $\theta$ Y component for L&RR);  1) For all patients, weak-moderate but significant correlation between balance testing & both rotation HRA, and SPNT; weak correlation between SPNT & HRA. For WAD-D, stronger correlation between LR HRE and SPNT; No correlations for WAD-ND.  2) An abnormal rotational HRA score had high +ve prediction value (88%) but low sensitivity (60%) and specificity (54%) to determine abnormality in balance and SPNT.  <u>Other</u> – Need to use all three measures to identify disturbances in postural control system.



Table 1. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Zito, 2006	HRA-SSA	Observation of KS Difference; Female Headache Patient subgroups (cervicogenic headache), migraine with aura vs Healthy Control group	<p>Patients (CHA): n=27; (mean age <math>\pm</math> SD, 25.3 <math>\pm</math> 3.9 yrs)</p> <p>Patients (Miigraine): n=25; ( mean age <math>\pm</math> SD, 22.9 <math>\pm</math> 3.5 yrs)</p> <p>Asymptomatic: n=25; (mean age <math>\pm</math> SD, 22.9.9 <math>\pm</math> 3.5 yrs)</p>	<p>'Cervicogenic Headache' (Sjaastad)</p> <p>Duration &gt; 3mo</p>	<sup>a</sup> Laser pointer	<p>As per Revel 1991 with following differences and uncertainties</p> <p>Procedure not described further but used 5 reps</p> <p>HRA Tasks - LR, RR, Ext, Flex <math>\Rightarrow</math> 0;</p>	<p><u>Absolute HRA-SSA</u></p> <p>Error component in 1° movement planes used for analysis (e.g. 0Y component for L&amp;RR)</p> <p>No significant difference in HRA-SSA between any group with all trial directions</p>
Teng, 2007	HRA-SSA  (Really assessed NHP)	<p>Observation of KS difference in 3 asymptomatic subject subgroups</p> <p>1) To determine whether a history of neck pain contributes to an abnormality in KS (middle aged adult groups)</p> <p>2) The effect of age on KS (young adult group and middle aged adults)</p>	<p><u>Gp1 Subjects</u> (young adult): n=20; (M:F, 11:9; mean age <math>\pm</math> SD, 21.9 <math>\pm</math> 3.9 yrs [range, 18 to 30])</p> <p><u>Gp2 Subjects</u> (Middle aged – NO history of NP): n=20; (M:F, 3:17; mean age <math>\pm</math> SD, 54.5 <math>\pm</math> 5.0 yrs [range, 45 to 65])</p> <p><u>Gp3 Subjects</u> (Middle aged –history of NP): n=20; (M:F, 6:14; mean age <math>\pm</math> SD, 58.8 <math>\pm</math> 5.7 yrs [range, 45 to 65])</p>	<p>All Current 'Asymptomatic' Subjects</p> <p>Gp1 Subjects: defined as 'no previous treatment for NP &amp; no current NP.</p> <p>Gp2 Subjects: no history of NP; no current NP</p> <p>Gp3 Subjects: history of chronic mild insidious onset, mechanical neck pain (&gt;6 months in past few years); no current NP.</p>	<sup>d</sup> Zebbris CMS70P	<p>Eyes Closed</p> <p>NHP <math>\rightarrow</math> Passive rotation of head (&lt; 35°/sec) to nSA target <math>\rightarrow</math> NHP <math>\rightarrow</math> nSA target.</p> <p>One rep of each task? - stated that head <u>not</u> repositioned between trials for each task.</p> <p>Measured in degrees</p> <p>HRA Tasks - RR, LR, LLF, RFL Ext &amp; Flex <math>\Rightarrow</math> 0; Order not given</p>	<p><u>RMS HRA-SSA (component in plane of movement?)</u></p> <p>1) No difference between the middle aged groups (although pain group generally larger errors)</p> <p>2) HRA worse for middle aged group (Gp2) of young adults (Grp1), with all tasks, particularly in sagittal plane (Flex/ext). The differences in this plane were statistically significant (<math>P &lt; 0.05</math>).</p> <p><u>Signed HRA-SSA (Overshoot/undershoot)</u></p> <p>1) No difference between the middle aged groups</p> <p>2) Overshoot for both middle aged group (Gp2) young adults (Grp1) with all tasks. This was obviously larger in Gp2 particularly in sagittal plane (Flex/ext). The differences in this plane were statistically significant (<math>P &lt; 0.05</math>).</p>

**Table 1. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Lark, 2007	HRA-SSA (Really assessed NHP)	Observation of KS Difference; Rugby players vs non-rugby sports people  (also measured full active c-spine ROM)	Rugby forwards: n=26 (all male; mean age $\pm$ SD, 26 $\pm$ 5 yrs)  Rugby backs: n=20 (all male; mean age $\pm$ SD, 24 $\pm$ 5 yrs)  Non-Rugby: n=14 (all male; mean age $\pm$ SD, 20 $\pm$ 7 yrs)	All groups with 'No Current Neck Injuries'  Rugby players: Division 1, II & premiership  Non-rugby players: sports people who had competed at regional or national standard in their sport.	<sup>a</sup> CROM	Eyes closed  NHP $\rightarrow$ Maximum rotation of head (2 sec hold) $\rightarrow$ NHP  Measured in degrees  HRA Tasks - Random order X2 reps of LR, RR, Flex, Ext, L & RLF	<u>Absolute HRA-SSA</u>  No difference between forwards and backs; therefore results combined. Combined rugby player group had higher errors in all movement directions (partic with Flex/Ext) except lateral flexion. The difference was only significant diff ( $P < 0.05$ ) for Ext trial. NB: larger SD for both groups with significant overlap.
Dunford, 2007	HRA-SSA (Really assessed NHP)	Observation of KS Difference in Rugby players; Forwards vs Backs  (also measured full active c-spine ROM)	Rugby forwards: n=18 (all male; mean age $\pm$ SD, 28 $\pm$ 8.8 yrs)  Rugby backs: n=17 (all male; mean age $\pm$ SD, 26 $\pm$ 4.4 yrs)	Rugby players: Two amateur clubs in Bournemouth & Bristol  Forwards: NP in previous 4 weeks – 15 (83%).  Backs: NP in previous 4 weeks – 7 (41%).  Prevalence Odds Ratio: 7.14	<sup>a</sup> Laser pointer  (mounted on cycling helmet)  90 cm from target	Eyes closed  Acclimatisation Movmnt; LR $\rightarrow$ NHP  NHP $\rightarrow$ Maximum rotation of head $\rightarrow$ NHP $\rightarrow$ eyes open & recentre on initial NHP $\rightarrow$ Maximum rotation of head $\rightarrow$ NHP  Measured in degrees  HRA Tasks - LR, RR; repeated x3	<u>Global HRA-SSA</u>  No significant difference between forwards and back in HRA.  6 of 18 forwards (33%) & 6 of 17 backs (35%) had repositioning error $> 4.5^\circ$

**Table 1. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Lee, 2007	HRA-SSA  (Really assessed NHP)  See also HRA-nSSA (Table 2)	Observation of KS difference in subject subgroups classified by pain characteristics.	Subjects: n=127; ( mean age $\pm$ SD, 41.8 $\pm$ 8.5 yrs)	'Subclinical' Neck pain'  Characterised in 3 ways <i>Pain Frequency:</i> Monthly, weekly, daily. <i>Pain Episode Duration:</i> <10 min, < 1hr, and > 1 hr. <i>Pain intensity:</i> Minimal, Mild, Moderate	<sup>d</sup> Zebris CMS70P	Eyes Closed  NHP $\rightarrow$ Movement of head to self-selected mid position $\rightarrow$ NHP 3 repetitions Measured in degrees  HRA Tasks – Flex , Ext, LR and RR $\Rightarrow$ 0 (random order of trials)	<u>Absolute HRA-SSA</u>  <i>Pain Frequency:</i> 1) Larger error for daily pain vs monthly for all trials – significant diff ( $P < 0.01$ ) for L & RR & Ext 2) Larger error for for weekly pain vs monthly for all but Ext trial – significant difference for L & RR, only.  Neither Pain Intensity or duration was found to have any consistent effect on repositioning errors  For HRA-SSA, significantly ( $P < 0.01$ ) larger error sizes were seen with increasing age
Demaille-Wlodyka, 2007	HRA-SSA  (Really assessed NHP)	Observational study  To determine the effect of age and gender on the ability of healthy asymptomatic subjects to reproduce the neutral zero head position.	<u>Asymptomatics</u> (total): n=232; age range, 15 to 79 yrs	'Healthy' Subjects  No current or prior history of musculoskeletal pain in neck or upper limbs, myopathy, dizziness or vertigo.  Group according to age;  15-24, 25-34, 35-44, 45-54, 55-64, >65 yrs	<sup>d</sup> Zebris CMS70P	Eyes closed,  HRA measured during ROM assessment Last plane in series: NHP $\rightarrow$ maximum rot (3 secs hold) $\rightarrow$ NHP  Subjects repositioned after RR trial.  10 repetitions  HRA Tasks – RR & LR $\Rightarrow$ 0 (same order for each subject?)	<u>Absolute HRA-SSA</u> (mean $\pm$ SD)  $\theta$ Y component (axial/horizontal rotation plane) used for analysis?  No effect of age or gender on HRA in horizontal plane. RR 4.23° $\pm$ 3.55 LR) 4.68° $\pm$ 4.05  Note ROM decreased with age ( $\therefore$ ROM not related to HRA?)

Table 1. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Treleaven, 2008	HRA-SSA (Really assessed NHP)	Observation of KS Difference; Patient Subgroups (unilateral vestib pathol & whiplash injury) vs Healthy Control group	<p>Patients (unilat vestib): n=20; (M:F, 11:9; mean age, 51yrs [range, 33 to 59])</p> <p>Patients (whiplash injury): n=20; (M:F, 5:15; mean age, 46.5yrs [range, 40 to 60])</p> <p>Asymptomatic: n=20; (M:F, 6:14; mean age, 49.5yrs [range, 43 to 59])</p>	<p>-Unilateral Vestibular Path: Acoustic Neuroma; 12 had the tumours removed (3 – 48 mo; mean – 15mo). Mean time since removal of tumour or diagnosis – 23 mo</p> <p>-Whiplash Injury patients (WAD): time since injury (mean) 17 mos (range, 4 to 36 mo); Primary complaint of dizziness or unsteadiness.</p>	°3-Space Fastrak	<p>As per Revel 1991 (see earlier) with following differences</p> <p>Order was randomized with 2 other tests; SPNT &amp; Postural stability</p> <p>Rotation of the head to comfortable limits</p> <p>3 repetitions</p> <p>HRA Tasks - LR, RR, &amp; Ext ⇒ 0; same order for each subject</p>	<p><u>Absolute HRA-SSA</u></p> <p>Both whiplash and vestibular patients showed a larger repositioning error compared with controls. These diffs were significant (<math>P &lt; 0.05</math>) RR &amp; Ext.</p> <p>Vestibular patients showed larger errors than WAD patients for L &amp; RR but these differences were not significant.</p>
Sjölander, 2008	HRA-SSA (Really assessed NHP)	Observation of KS Difference; Patient subgroups (insidious onset & whiplash injury) vs Healthy Control group	<p>Patients (insidious onset): n=9; (M:F, 0:9; mean age ± SD, 40 ± 9 yrs)</p> <p>Patients (whiplash injury): n=7; (M:F, 2:7; mean age ± SD, 45 ± 11 yrs)</p> <p>Asymptomatic: n=16; (M:F, 3:13; mean age ± SD, 41 ± 9 yrs)</p>	<p>'Chronic mechanical Cervical Pain': &gt; 6 mths duration</p> <p>-Insidious Onset patients: Pain duration (mean ± SD); 97 ± 68 mths; current severity (VAS - mm), 52 ± 26; NDI of 37 ± 11; vertigo/unsteadiness: 2/1</p> <p>- Whiplash Injury patients (WAD II-III): Pain duration (mean ± SD); 76 ± 84 mths; current severity (VAS-mm), 45 ± 19; NDI of 44 ± 23. vertigo/unsteadiness: 4/5</p>	°3-Space Fastrak	<p>Eyes closed</p> <p>Standing</p> <p>Order of HRA was randomized with 2 other tests; Standing balance &amp; cervical stability</p> <p>NHP → rotation of head as fast &amp; far as possible → final position of previous rotation, short rest</p> <p>8 repetitions</p> <p>HRA Tasks – RR &amp; LR ⇒ 0 (same order for each subject?)</p>	<p><u>Signed (CE) HRA-SSA</u></p> <p>θY component (axial/horizontal rotation plane) used for analysis</p> <p>Slightly larger CE in patient groups vs controls but not significant</p> <p><u>Variable (VE) HRA-SSA</u></p> <p>Larger VE in L &amp; RR for patient groups vs controls (WAD &gt; Insidious).</p> <p>For RR, a significant difference (<math>P &lt; 0.01</math>) was seen for insidious and WAD vs controls; No significant difference between patient groups: ROM was also a significant covariate (<math>P &lt; 0.05</math>)</p> <p>For LR, significant difference (<math>P &lt; 0.01</math>) between control &amp; WAD group. ROM was not a significant covariate</p>

**Table 1. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
De Hertogh, 2008	HRA-SSA (Really assessed NHP)	Observation of KS Difference; Patient vs Healthy Control group	<u>Patients</u> : n=10; (M:F, 1:9; mean age $\pm$ SD 41 $\pm$ 16 yrs  <u>Asymptomatics</u> : n=24; (M:F, 2:22; mean age $\pm$ SD 34 $\pm$ 12 yrs)	'Cervicogenic Headache' (Sjaastad)  Duration > 3mo	Flock of Birds electromagnetic tracking device.	As per Revel 1991 (see earlier) with following differences  Not clear if any rest between trials  Pts asked to say 'yes' when reached NHP  HRA Tasks - RR, LR, 5 mins rest, Flex & Ext $\Rightarrow$ 0; same order for each subject	<u>Absolute HRA-SSA</u> Error component in 1° movement planes used for analysis (e.g. $\theta Y$ component for L&RR)  No significant difference between patient and control subjects.

**Abbreviations:**

$\theta X$  – angle of rotation around X-axis (flex/ext);  $\theta Y$  – angle of rotation around Y-axis (L & R rotation);  $\theta Z$  – angle of rotation around Z-axis (L & R lateral flexion).

Ext $\Rightarrow$  0 & Flex $\Rightarrow$  0: Repositioning from extension and flexion to reference zero (SSA or NHP)

HRA-SSA : head repositioning accuracy to subjective straight ahead.

KS: Kinaesthetic Sensibility

LR – Left rotation; RR – Right rotation; Ext – Extension; Flex – Flexion

LR $\Rightarrow$  0 & RR $\Rightarrow$  0: Repositioning from left and right rotation to reference zero (SSA or NHP)

NDI: Neck Disability Index

NHP: Natural or neutral head position

nSA: non-straight ahead

OA: Osteoarthritis

PSFS: Patient-Specific Functional Scale

ROC: Receiver operating curve

ROM: Range of motion

Rx: Treatment

SMT: Spinal manipulative therapy

VAS: Visual Analogue Scale

WAD: Whiplash Associated Disorder (grades I-IV)

<sup>†</sup> Studies available before commencing this thesis

<sup>a</sup>laser pointer technique introduced by Revel *et al* (1991) testing the ability of blindfolded subjects to accurately relocate the head, to a subjective 'straight-ahead' position, after a maximal active movement of the head in the horizontal or vertical plane.

<sup>b</sup>CA 6000 Spine Motion Analyser – a link arm connecting six high precision potentiometers positioned to measure motion in the 3 cardinal planes of motion. Device is attached to patient with two harnesses at end of link arm.

<sup>c</sup>3-Space Fastrak – an electromagnetic measuring instrument which tracks the positions of sensors relative to a source in three dimensions.

<sup>d</sup>Zebris CMS20/70P – a method of measurement based on the determination of the spatial coordinates around three orthogonal axes of miniature ultrasound (US) transmitters relative to a fixed system of three microphones.

<sup>e</sup>cervical range-of-motion device – a plastic device consisting of a magnetic yoke, resting on the shoulders, and a plastic headpiece with three goniometers positioned to measure the three cardinal planes of movement.

**Table 2.** Investigations of cervicocephalic kinaesthetic sensibility utilizing repositioning and positioning tasks to non-neutral set points with the cervical range of motion.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
<sup>†</sup> Loudon, 1997	HRA-nSA	Observation of KS Difference; Patient vs Healthy Control group	Patients: n=11; (M:F, 2:9; mean age $\pm$ SD, 42 $\pm$ 8.7 yrs [range, 28 to 57])  Controls: n=11; (M:F, 2:9; mean age $\pm$ SD, 43 $\pm$ 1.3 yrs [range, 28 to 57])	'Whiplash Injury'  1-3 whiplash injuries within 2 years of testing date  At least 3 months since last injury  Neck Pain & limited ROM – details not given.	<sup>o</sup> CROM	Eyes closed NHP $\rightarrow$ Passive rotation of head nSSA target $\rightarrow$ NHP $\rightarrow$ nSSA target  3 reps ('trials') for each HRA task (not repositioned between trials)  Measured in degrees HRA Tasks – LR & RR $\Rightarrow$ 30° & 50° horizontal rotation; LLF & RLF $\Rightarrow$ 20° lateral flexion (order & randomization?)	Error component 1° movement planes (e.g. $\theta$ Y component for L&RR) used for analyses <u>Absolute HRA-SSA</u>  The average absolute error was 5.0° for patients and 1.75° for controls; not stated if 50° tasks were less accurate than 30°.  <u>Constant HRA-SSA (Overshoot/undershoot)</u>  Overshoot tendency for both groups with all tasks but greater with patient group (except RR 50°) Significant diff ( $P < 0.05$ ) between patients and controls for all tasks (except RR 50°) Also a much larger variation in scores (SD) reflecting larger absolute error.  Within groups, significantly larger overshoot between LR 30° task and both lateral flexion tasks. No diff in overshoot between tasks for control group <u>Other</u> – Authors reported that whiplash group demonstrated a NHP at the start of task other than (0,0,0) – what they perceived as neutral was actually tilted or rotated wrt trunk.
Dumas, 2001	HRA-nSA	Observation of KS Difference; Headache Patient subgroups (cervicogenic headache with trauma [CHA-T], CHA insidious in onset [CHA-I] migraine) vs Healthy Control group	Patients (CHA-T): n=20; (mean age $\pm$ SD, 45.3 $\pm$ 11.4 yrs)  Patients (CHA-I): n=24; (mean age $\pm$ SD, 44.3 $\pm$ 11.9 yrs)  Patients (Migraine): n=16; (mean age $\pm$ SD, 39 $\pm$ 12.5 yrs)  Controls: n=17; (mean age $\pm$ SD, 43 $\pm$ 14.1 yrs)	'Cervicogenic Headache' (Sjaastad)  Duration > 6mo	<sup>o</sup> CROM	As per Loudon 1997	Error component 1° movement planes (e.g. $\theta$ Y component for L&RR) used for analyses <u>Absolute HRA-SSA</u>  No significant difference found between the four groups.  The average absolute error was 3.7° - for 4.5° patients and 3.8° for controls
Kristjansson, 2001	HRA-nSA	Test-Retest 'Reliability' (Study primarily aimed at assessing reliability of 5 different KS tests in healthy subjects – reliability results reported in Chapter 5)	Subjects: n=19; (7 men and 12 women, mean age $\pm$ 2SD, 35.1 $\pm$ 10 yrs).	'Healthy' Subjects  No current or prior history of musculoskeletal pain in neck or upper limbs.	<sup>o</sup> 3-Space Fastrak	Vision Occluded  3 repetitions Measured in degrees Tasks; RR & LR  <u>Test 2</u> - 30° L or RR of head (examiner positioned) $\rightarrow$ SSA $\rightarrow$ 30° L or RR <u>Test 3</u> - 30° L or RR of trunk (examiner positioned) $\rightarrow$ SSA $\rightarrow$ 30° L or RR	<u>Absolute HRA-SSA</u> (mean $\pm$ SD) $\theta$ Y component (axial/horizontal rotation plane) used for analysis Test 2) 5.85° $\pm$ 3.81 Test 3) 5.96° $\pm$ 4.65

Table 2. Contd.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Kristjansson, 2003	HRA-nSA	Observation of KS Difference; Patient subgroups (insidious onset & whiplash injury) vs Healthy Control group	<p>Patients (insidious onset): n=20; (M:F, 11:9; mean age <math>\pm</math> SD, 30.1 <math>\pm</math> 9.1 yrs)</p> <p>Patients (whiplash injury): n=22; (M:F, 11:11; mean age <math>\pm</math> SD, 33.4 <math>\pm</math> 10.6 yrs)</p> <p>Asymptomatic: n=21; (M:F, 10:11; mean age <math>\pm</math> SD, 26.9 <math>\pm</math> 6.4 yrs)</p>	<p>'Chronic mechanical Cervical Pain': 3 &amp; 48 mths duration</p> <p>-Insidious Onset patients: Pain duration (mean <math>\pm</math> SD); 28.6 <math>\pm</math> 15.5 mths; current severity (VAS), 1.82 <math>\pm</math> 2.0.</p> <p>-Whiplash Injury patients: Pain duration (mean <math>\pm</math> SD); 21.9 <math>\pm</math> 12.5 mths; current severity (VAS), 3.37 <math>\pm</math> 2.8.</p>	$\infty$ 3-Space Fastrak	<p>As per Kristjansson, 2001 (see above) with following change:</p> <p>Random order of testing</p>	<p><u>Absolute HRA-SSA</u> (mean <math>\pm</math> SD)</p> <p>0Y component (axial/horizontal rotation plane) used for analysis</p> <p>The patient HRA was slightly worse than the asymptomatic groups but &lt; 1.0° in all but one instance with a wider variation for patient groups. No statistically significant difference was seen between the groups. The HRA scores were generally worse for all groups with test 3 (trunk rotation); for controls this was different to Kristjansson, 2001 (see above)</p> <p>NB: The whiplash Injury patients had significantly higher pain and disability scores than insidious group but the insidious group exhibited slightly worse HRA in both tests.</p>
Armstrong, 2005	HRA-nSA	<p>1) Observation of KS Difference; Patient vs Asymptomatic Control group</p> <p>2) Effect of Deep neck Flexor(DNF) activation on KS</p> <p>4 grps: Symptomatic + DNF(SDNF); symptomatic no DNF (SNDNF). Asympt + DNF(ADNF); Asympt no DNF (ANDNF).</p>	<p><u>Patients</u> (total): n=23; (M:F, 8:15; mean age <math>\pm</math> SD, 41.2 <math>\pm</math> 11.9 yrs [range, 20.4 to 67.9]))</p> <p><u>Asymptomatics</u> (total): n=23; (M:F, 10:13; mean age <math>\pm</math> SD, 33.9 <math>\pm</math> 12.1 yrs[range, 19.2 to 62.5])</p>	<p>'Whiplash Syndrome'</p> <p>WAD II &amp; III (87% graded as WAD II)</p> <p>Duration of symptoms (mean <math>\pm</math> SD); 31 <math>\pm</math> 32 months.</p> <p>NDI score (mean <math>\pm</math> SD); 12 <math>\pm</math> 5.5</p> <p>PSFS score (mean <math>\pm</math> SD); 4.7 <math>\pm</math> 1.6</p>	$\infty$ 3-Space Fastrak*	<p>Vision Occluded</p> <p>NHP <math>\rightarrow</math> random position within mid-range of rotation or flex or inner range of ext ('target position') <math>\rightarrow</math> NHP <math>\rightarrow</math> 'target position'</p> <p>3 trials in each direction of motion</p> <p>Measured in degrees (*at each receptor site)</p> <p>HRA Tasks – LR, RR, Flex and Ext (not clear on order within or between subjects)</p>	<p><u>Absolute HRA</u> (mean of all specific movements 'pooled' together)</p> <p>1) No significant difference in HRA between whiplash &amp; healthy subjects (data 'pooled' for intervention results)</p> <p>2) No significant difference in HRA between DNF and no DNF groups</p> <p><u>Constant HRA (Overshoot/undershoot)</u></p> <p>1) No significant difference in overshoot/undershoot characteristics between groups or tasks</p> <p>2) No significant difference in overshoot/undershoot characteristics between DNF and no DNF groups</p> <p>NB: Irrespective of group, type of task or intervention, 65-80% of trials resulted in overshoot (all &lt; 1.0°?)</p> <p>Other: No significant correlation between absolute HRA &amp; NDI, PSFS, Active ROM; pain increased by 15% during the trials but no significant correlation between absolute HRA and pain severity and duration</p>
Lee, 2005	HRA-nSA	<p>Observational study</p> <p>To determine the ability of subclinical NP subjects to reproduce the neutral zero head position.</p>	<p>Subjects: n=81; ( mean age <math>\pm</math> SD, 23.2 <math>\pm</math> 3.3 yrs)</p>	<p>'Subclinical' Neck pain</p> <p>Defined as: no history of neck, upper back or spinal problems that resulted in a restriction of normal activity or time lost from work.</p> <p>Subjects excluded if had med attention for neck or related complaints within last 6 mo</p> <p>NP characteristics quantified using SFMPQ, FRI, NPDS, NDI – used to classify subjects into 3 pain groups.</p> <p>- Never/infrequent (&lt; 6/yr)</p> <p>- Monthly (1-3/mth)</p> <p>- Weekly (1/wk to daily)</p>	Purpose engineered head position measurement apparatus	<p>Vision maintained straight ahead</p> <p>Cheek on fixed plate <math>\rightarrow</math> Movement of head to moveable plate (target angle) <math>\rightarrow</math> identify which of five target positions had been contacted</p> <p>Each testing position presented 10 times in random order</p> <p>Subjects first shown the five different locations and given three practises <u>with feedback</u></p> <p>KS tasks – LR, RR and hd retract movements.</p>	<p><u>Just Noticeable Difference (JND) for Movement Extent</u></p> <p>(JND signifies movement sensitivity; a lower JND value represents better discrimination)</p> <p>Overall with movement direction data combined, a trend for lower JND (better discrimination) was seen from the never/infrequent to the combined data of the more frequent pain groups.</p> <p>With combined movement directions, the weekly pain group were significantly more sensitive than the monthly pain group (<math>P &lt; 0.05</math>).</p> <p>When movements separated to rotation and retraction, the never/infreq group were had far lower JND for retraction than rotation (hence no diff with the combined movements for pain and never/infrequent movements).</p>

**Table 2. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results															
Feipel, 2006	HRA-nSA	Observation of KS Difference; Patient vs 'Healthy' Control group	<p>Patients: n=29; (62% female; mean age <math>\pm</math> SD, 37 <math>\pm</math> 14 yrs)</p> <p>Controls: n=26; (54% female; mean age <math>\pm</math> SD, 35 <math>\pm</math> 11 yrs)</p>	<p>'Whiplash Syndrome'</p> <p>WAD I-III (traffic or sports injury); 29% graded as WAD I-II</p> <p>Duration (mean time from accident <math>\pm</math> SD); 31 <math>\pm</math> 32 months.</p>	<sup>b</sup> CA 6000 Spine Motion Analyser	<p>Vision Occluded</p> <p>NHP <math>\rightarrow</math> Active positioning of head to nSA position (guided by investigator) 3 secs <math>\rightarrow</math> NHP <math>\rightarrow</math> active nSA angle</p> <p>4 reps of angle</p> <p>Measured in degrees</p> <p>Tasks – Pure L &amp; RR 50° &amp; combined Rot (50°) and lateral flexion (20°).</p>	<p>Significant (<math>P &lt; 0.05</math>) difference (WAD larger error) in HRA for WAD grp vs controls. However differences were small</p> <p>Errors were larger for the nSA tasks compared with the SSA task (table 1).</p> <p>Error was largest in plane of movement.</p>															
McNair, 2006	HRA-nSA	<p>Case Report</p> <p>Quantify KS before and after specific cervical mobilization techniques.</p>	Patient: 44 year old male	<p>Acute Idiopathic neck pain</p> <p>Duration: 1 day</p> <p>Left sided mid cervical pain with radiation to L upper trap</p> <p>Pain severity: 5-6/10</p> <p>Neck stiffness &amp; reduced ROM</p>	<sup>d</sup> Zebis CMS70P	<p>Vision Occluded</p> <p>NHP <math>\rightarrow</math> Arbitrary position in rotation or flex/ext ('target position') <math>\rightarrow</math> NHP <math>\rightarrow</math> 'target position'</p> <p>2 trials in each direction of motion</p> <p>Measured in degrees</p> <p>HRA Tasks – LR, RR, Flex and Ext (not clear on order within or between subjects)</p>	<p><u>Absolute HRA-SSA (component in plane of movement?)</u></p> <table border="1"> <thead> <tr> <th></th> <th>Before Rx</th> <th>After Rx</th> </tr> </thead> <tbody> <tr> <td>RR</td> <td>4°</td> <td>4°</td> </tr> <tr> <td>LR</td> <td>2°</td> <td>3°</td> </tr> <tr> <td>Flex</td> <td>2°</td> <td>2°</td> </tr> <tr> <td>Ext</td> <td>1°</td> <td>1°</td> </tr> </tbody> </table> <p>Other: Increased ROM in Flex (55%), Ext (35%), LR (56%) &amp; LLF (22%) – for rot and LF, these were the least restricted before Rx; Reduction in average pain to 2/10 (worst being 2.5/10)</p>		Before Rx	After Rx	RR	4°	4°	LR	2°	3°	Flex	2°	2°	Ext	1°	1°
	Before Rx	After Rx																				
RR	4°	4°																				
LR	2°	3°																				
Flex	2°	2°																				
Ext	1°	1°																				



**Table 2. Contd.**

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Teng, 2007	HRA-nSA	<p>Observation of KS difference in 3 asymptomatic subject subgroups</p> <p>1) To determine whether a history of neck pain contributes to an abnormality in KS (middle aged adult groups)</p> <p>2) The effect of age on KS (young adult group and middle aged adults)</p>	<p><u>Gp1 Subjects</u> (young adult): n=20; (M:F, 11:9; mean age <math>\pm</math> SD, 21.9 <math>\pm</math> 3.9 yrs [range, 18 to 30])</p> <p><u>Gp2 Subjects</u> (Middle aged – NO history of NP): n=20; (M:F, 3:17; mean age <math>\pm</math> SD, 54.5 <math>\pm</math> 5.0 yrs [range, 45 to 65])</p> <p><u>Gp3 Subjects</u> (Middle aged –history of NP): n=20; (M:F, 6:14; mean age <math>\pm</math> SD, 58.8 <math>\pm</math> 5.7 yrs [range, 45 to 65])</p>	<p>All Current 'Asymptomatic' Subjects</p> <p>Gp1 Subjects: defined as 'no previous treatment for NP &amp; no current NP.</p> <p>Gp2 Subjects: no history of NP; no current NP</p> <p>Gp3 Subjects: history of chronic mild insidious onset, mechanical neck pain (&gt;6 months in past few years); no current NP.</p>	<sup>d</sup> Zebris CMS70P	<p>Eyes Closed</p> <p>NHP <math>\rightarrow</math> Passive rotation of head (&lt; 35°/sec) to nSSA target <math>\rightarrow</math> NHP <math>\rightarrow</math> <u>nSSA target</u>.</p> <p>One rep of each task? - stated that head not repositioned between trials for each task.</p> <p>Measured in degrees</p> <p>HRA Tasks - RR, LR, LLF, RFL Ext &amp; Flex <math>\Rightarrow</math> 0; Order not given</p>	<p><u>RMS HRA-SSA (component in plane of movement?)</u></p> <p>1) No difference between the middle aged groups</p> <p>2) HRA worse for middle aged group (Gp2) of young adults (Grp1), with all tasks, particularly in horizontal plane (L &amp; RR). No statistically significant differences were seen</p> <p><u>Constant HRA-SSA (Overshoot/undershoot)</u></p> <p>1) No difference between the middle aged groups</p> <p>2) Overshoot for both middle aged group (Gp2) young adults (Grp1) with most tasks (2 undershoots in control grp). This was obviously larger in Gp2 particularly in horizontal plane (L &amp; RR). No statistically significant differences were seen.</p>
Lee, 2007	HRA-nSA	<p>Observation of KS difference in subject subgroups classified by pain characteristics.</p>	<p>Subjects: n=127; ( mean age <math>\pm</math> SD, 41.8 <math>\pm</math> 8.5 yrs)</p>	<p>'Subclinical' Neck pain'</p> <p>Characterised in 3 ways</p> <p><i>Pain Frequency:</i> Monthly, weekly, daily.</p> <p><i>Pain Episode Duration:</i> &lt;10 min, &lt; 1hr, and &gt; 1 hr.</p> <p><i>Pain intensity:</i> Minimal, Mild, Moderate</p>	<sup>d</sup> Zebris CMS70P	<p>Eyes Closed</p> <p>NHP <math>\rightarrow</math> Movement of head to self-selected mid position (50% ROM)</p> <p>2 repetitions</p> <p>Measured in degrees</p> <p>HRA Tasks – Flex , Ext, LR and RR <math>\Rightarrow</math> 0 (random order of trials)</p>	<p><u>Absolute HRA-nSA</u></p> <p>Errors smaller for nSA compared with SSA (table1)</p> <p><i>Pain Frequency:</i></p> <p>1) Larger error for daily pain vs monthly for all trials – significant diff (<math>P &lt; 0.05</math>) for LR</p> <p>2) Larger error for weekly pain vs monthly for all but LR trial – No significant diff.</p> <p>Neither Pain Intensity or duration was found to have any consistent effect on repositioning errors</p> <p>For HRA-nSA, significantly (<math>P &lt; 0.01</math>) larger error sizes for females.</p>
<p><i>Abbreviations:</i> See table 1</p>							

**Table 3.** Investigations of cervicocephalic kinaesthetic sensibility utilizing dynamic head movement control tasks.

Principal Author	KS Task	Study Design/objectives	Number of subjects & demographics	Clinical & Pathological Characteristics of Patient Subjects	Measurement Method for KS	Measurement Procedure for KS	Summary of Main Reported KS Results
Kristjansson, 2001	Head movement control	Test-Retest 'Reliability' (Study primarily aimed at assessing reliability of 5 different KS tests in healthy subjects – reliability results reported in Chapter 5)	Subjects: n=19; (7 men and 12 women, mean age $\pm$ 2SD, 35.1 $\pm$ 10 yrs).	'Healthy' Subjects No current or prior history of musculoskeletal pain in neck or upper limbs.	$\epsilon$ 3-Space Fastrak	Eyes closed Measured in degrees  <u>Test 5</u> x3 Fig 8 movements of head (without stopping) through the NHP (∴ 5 passes through NHP)	<u>Absolute HRA-SSA</u> (mean $\pm$ SD) 0Y component (axial/horizontal rotation plane) used for analysis Test 5) 4.82° $\pm$ 3.08
Kristjansson, 2003	Head movement control	Cross-sectional Observation of KS Difference; Patient subgroups (insidious onset & whiplash injury) vs Healthy Control group	Patients (insidious onset): n=20; (M:F, 11:9; mean age $\pm$ SD, 30.1 $\pm$ 9.1 yrs)  Patients (whiplash injury): n=22; (M:F, 11:11; mean age $\pm$ SD, 33.4 $\pm$ 10.6 yrs)  Asymptomatic: n=21; (M:F, 10:11; mean age $\pm$ SD, 26.9 $\pm$ 6.4 yrs)	'Chronic mechanical Cervical Pain': 3 & 48 mths duration  -Insidious Onset patients: Pain duration (mean $\pm$ SD); 28.6 $\pm$ 15.5 mths; current pain severity (VAS), 1.82 $\pm$ 2.0.  -Whiplash Injury patients: Pain duration (mean $\pm$ SD); 21.9 $\pm$ 12.5 mths; current severity (VAS), 3.37 $\pm$ 2.8.	$\epsilon$ 3-Space Fastrak	As per Kristjansson, 2001 (see above) with following change:  Random order of testing	<u>Absolute HRA-SSA</u> (mean $\pm$ SD) 0Y component (axial/horizontal rotation plane) used for analysis  The whiplash patient HRA was slightly worse than the other groups but with a wider variation (SD). No statistically significant difference was seen between the groups.  NB: The insidious NP group performed better than the control group (5.82 $\pm$ 3.08 vs 6.86 $\pm$ 3.94)
Kristjansson, 2004	Head movement control	Cross-sectional Observation of KS Difference; Patient vs Asymptomatic Control group	All subjects Female Mean age ( $\pm$ SD), 30.0 yrs ( $\pm$ 8.8) Patients: n=20;  Asymptomatic: n=20	'Chronic Whiplash Syndrome' (WAD grades I & II) > 6 mths  Current pain severity (VAS), 46.8 $\pm$ 21.0.  Northwick Park Neck Pain Disability Index of 45% $\pm$ 14%.	$\epsilon$ 3-Space Fastrak	Eyes open  Practice movement pattern performed by all subjects  Task – match movement of forehead referenced cursor to computer generated cursor path (pattern)  3 repetitions (trials)  3 tasks (patterns A-C)  Random order of test patterns and trials	<u>Absolute Error</u> Not clear which orthogonal component used Significantly ( $P < 0.05$ ) greater errors with all movement patterns for the WAD group vs controls. NB: CI's suggest that can discriminate between the 2 groups?

Abbreviations:

See table 1

## Chapter 2

---

# HEAD REPOSITIONING ACCURACY USING THE LASER POINTER METHOD IN PATIENTS WITH CHRONIC INSIDIOUS ONSET NECK PAIN

---

### 2.1 INTRODUCTION

From many clinical perspectives, patients with chronic mechanical neck pain represent a heterogeneous group. One obvious subclassification stated earlier relates to the nature of onset for the cervical spine complaint namely; insidious (non-traumatic) and ‘whiplash’ injury. As also highlighted in Chapter 1, prior to commencing the current study, several studies had found diminished cervicocephalic kinaesthesia in chronic neck pain patients where the cause was either not stated (Revel et al., 1991), of mixed onset (Revel et al., 1994) or a cervical ‘whiplash’ injury. (Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998). All these investigators used the HRA-SSA laser pointer technique introduced by Revel et al., (1991). By using a different measuring technique, a similar loss of kinaesthetic sensibility was demonstrated in chronic ‘whiplash’ patients when asked to relocate the head to various rotation and side bending positions (Loudon et al., 1997).

Cervicocephalic kinaesthetic sensibility had not been investigated specifically in patients who had no history of cervical spine injury from trauma such as ‘whiplash’. It is of interest to know if these patients also exhibit deficits in cervicocephalic sensibility or whether it is only seen with ‘whiplash’ patients. The fact that only 28% of patients in the study involving a mixed onset group (Revel et al., 1994) were reported to have chronic neck pain resulting from ‘minor’ trauma suggested that similar deficits in repositioning to SSA (or NHP) may exist with neck pain of an insidious onset. A more recent study also involving a mixed onset group (61% having “past” whiplash) supports this position with a clear impairment in HRA-NHP compared with asymptomatic control subjects using the laser pointer method (Humphreys and Irgens, 2002). Four studies subsequent

to this have also investigated aspects of cervicocephalic KS in patients with neck pain of insidious onset (Kristjansson et al., 2003; Palmgren et al., 2006; Teng, Chai, Lai and Wang, 2007; Sjolander et al., 2008) with mixed results. Using a similar protocol to that introduced by Revel et al., (1991) but with 3-D motion analysis equipment, Kristjansson et al., (2003) and Sjolander et al., (2008) found impaired head repositioning accuracy to NHP in patients with chronic insidious onset neck pain. However, Sjolander et al., (2008) only found a significant difference for the right rotation trial although the errors were greater with the left rotation for the patient groups compared with asymptomatic control subjects. Using the laser pointer but a different measurement protocol, improvements in HRA to an investigator determined NHP were seen in a cohort of chronic pain patients after a mixed program of active and passive cervical spine therapies (Palmgren et al., 2006). However, there is an assumption here that the patients exhibited a deficit in KS prior to treatment as there was no pre-treatment comparison with healthy controls.

Teng, Chai, Lia and Wang (2007) did not find any differences between a group of middle aged subjects with a history of neck pain and healthy control subjects. Once again, a variation in testing protocol and differences in measurement equipment and group characteristics are all possible reasons for the difference in results compared with earlier investigations. In particular, the patient group studied by Teng et al., (2007) were asymptomatic at the time of testing so perhaps were best classified as subclinical or chronic recurrent neck pain patients.

The aim of this preliminary study was to compare head repositioning accuracy (subjective 'straight-ahead' [SSA] of head on trunk) in patients with chronic, insidious onset cervical spine pain to an age and gender matched control group. The study would also serve as a preliminary exercise on the practical feasibility of future, larger scale comparative studies using the Anglo-European College of Chiropractic (AECC) clinic and to gain practical experience utilising the HRA-SSA procedures with controls subjects and patients.

## **2.2 METHODS**

### **2.2.1 Study Setting and Design**

This study took place in the outpatient clinic at the AECC, Bournemouth, UK. A two-group observational cohort design was used. Completion of questionnaires and all measurement procedures were conducted in one of the radiography rooms. The same room was used on each occasion.

### **2.2.2 Subject Recruitment and Selection**

Male and female patients were selected from all patients presenting for the first time at the AECC clinic over a six-week period. Throughout this time, all new patients completed a simple questionnaire as part of the inclusion/exclusion procedure. On daily review of these first stage questionnaires, the clinical records were reviewed in detail by an experienced member of the chiropractic faculty (the investigator), for those patients who provisionally met the inclusion criteria, indicated a willingness to participate and consented for their records to be inspected. Those subjects who met the inclusion/exclusion criteria (Table 4) were contacted by phone and invited to participate after a brief explanation of the nature of the study, the possible risks of participation and commitment required. Patients willing to take part were given an appointment with the investigator prior to their first treatment visit. At this measurement session the patients were given further verbal and written information about the study and asked to read and sign a consent form.

Control subjects were recruited from AECC staff, faculty and students. To be considered for inclusion, the subjects must have been aged 18 to 55 years; no history of 'whiplash' or other cervical spine injury/pain; no history of dizziness, vertigo or persistent or frequent headaches; no current treatment for any other musculoskeletal complaint and no systemic disease or any of the conditions listed under the exclusion criteria in Table 4. Eligible control subjects were finally selected by age and gender, to ensure a similar distribution to the patient group.

**Table 4.** Inclusion & Exclusion criteria

---

Inclusion

1. Age 18-55.
2. Males and females.
3. Continuous neck pain of more than 7 weeks duration.
4. Neck pain as main presenting complaint.

Exclusion

1. Onset of presenting neck pain episode following trauma (e.g., 'whiplash').
2. History of cervical injury of trauma since the onset of neck pain episode/s.
3. Prior history of cervical injury or trauma.
4. Cervical Radiculopathy &/or Myelopathy
5. Inflammatory Arthritis involving C-spine
6. Tumour or infection involving C-spine
7. Vertebrobasilar Artery Insufficiency
8. Neurological disease such as MS, MN, Parkinson's, Syringomyelia etc.
9. History of Dizziness
10. Known congenital anomalies involving the C-spine.
11. Systemic disease such as Diabetes Mellitus.

---

*Abbreviations:*

C-Spine – Cervical Spine

MS – Multiple Sclerosis

MND – Motor Neurone Disease

---

### 2.2.3 Outcome Measures and Equipment

***Clinical characteristics of the neck pain group.*** The main variables collected for each patient were the current intensity of pain using an 11-point numerical rating scale (NRS), duration and evolution of pain, and the unilateral or bilateral localisation of pain using pain drawings. Other psychometric data were gathered using the Bournemouth Questionnaire (Appendix I), with modifications for neck pain patients (Bolton and Breen, 1999; Bolton and Humphreys, 2002).

***Active range of cervical motion.*** A cervical range-of-motion device (CROM; Performance Attainment Associates, Roseville, Minnesota) was used to assess cervical motion in the transverse (horizontal rotation), sagittal (flexion-extension) and frontal (lateral-bending) planes (Figure 3). This device has been shown in several studies to have acceptable *reliability* in measuring cervical range-of-motion (ROM) (Capuano-

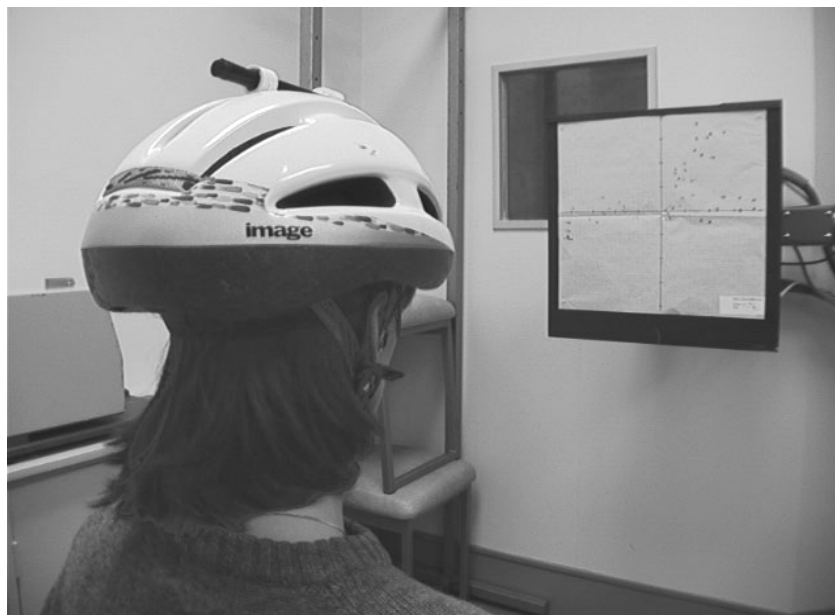
Pucci, et al., 1991; Youdas, Carey and Garrett, 1991). The equipment consists of a magnetic yoke, resting on the shoulders, and a plastic headpiece with three goniometers positioned to measure the three cardinal planes of movement (Figure 3). The transverse plane measurement involves a compass goniometer and the magnetic yoke. Measurement of sagittal and frontal plane motion utilises gravity goniometers. The same CROM instrument was used throughout the study.



**Figure 3.** Cervical range of motion device (CROM)

***Kinaesthetic sensibility test.*** Head repositioning accuracy to subjective straight ahead (HRA-SSA) was measured using a simple clinical technique adapted from the method first described by Revel et al., (1991). This involved the subject being blindfolded and wearing a lightweight cycling helmet (Bell Image™; Bell Sports Inc®, Rantoul, Illinois) with a laser pointer mounted on the top (total weight = 375g) (Figure 4). A square target holder was constructed (40 x 40cm) which slotted into the filter holder on the front of an x-ray tube. Targets were made from 40 x 40cm sheets of graph paper, ruled with 1mm resolution gridlines. Horizontal (X) and vertical (Y) axes were drawn on the paper so

that the axes intersected at the midpoint, dividing the paper into four quadrants. The point of intersection was used as the reference zero (0,0) position on the target. The mobility of the tube in X, Y and Z directions enabled the target's zero point to be accurately centred to the subject's reference head position.



**Figure 4.** Experimental apparatus and testing procedure used for evaluation of head repositioning accuracy.

#### **2.2.4 Measurement protocol**

The investigator conducted all measurement procedures. On arrival, the patients were asked to complete the questionnaires. As the control group were recruited a few weeks prior to the measurement session, they were checked again with respect to their inclusion/exclusion criteria. For both the CROM instrument and HRA measurements, the subjects were seated in a chair with a backrest for the lumbar and lower thoracic regions only. They were asked to sit as far back into the chair as possible, let their arms hang down by their sides and place the rear of their heels against specific foot markings



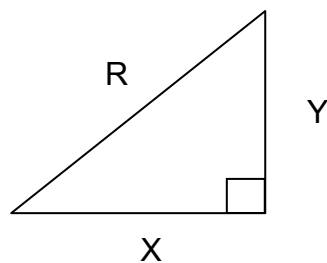
on the floor. They were told that the target would be straight ahead of them but would be fitted and moved into position only after they were blindfolded. The CROM device was then fitted to the head. The neutral position for the half-cycle ROM measurements was attained by asking the subjects to assume a 'straight ahead' posture with the head. Motion was measured in the following order for all patients and subjects; left rotation (LR), right rotation (RR), extension (Ext), forward flexion (Flex) left then right lateral flexion (LLF, RLF). The measurement routine was repeated three times and a mean angle (degrees) calculated for each motion direction. In the patient group, movement directions that were painful were also recorded. The CROM instrument was then removed.

The subjects next had the HRA testing procedure and objectives explained to them one more time. Both groups had a similar explanation delivered in a similar manner. Having completed this, all subjects were blindfolded using the sleeping mask for the remainder of the procedure and were instructed to keep their eyes closed behind the mask. The cycling helmet was then fitted and firmly secured to the head using the chinstrap. After the target was positioned, the room lights were dimmed and the subject was asked to find what they perceived as a 'straight ahead' position with their head. They were instructed to memorise this position and then relocate their head back to this position after one near-maximal amplitude extension movement of the head. The subject was told that this was the reference zero position and that they were to try to relocate back to this position as accurately as possible after each movement. The target was then moved so that the laser pointer's light beam projected on the zero of the target. After a few seconds concentration on this reference position, the subject was instructed to perform a near-maximal rotation of the head to the left (LR) and then immediately try to relocate back to the reference zero position as accurately as possible. No speed instruction was given. The point where the light beam stopped on the target was marked by the investigator with a dot from a pen and labelled according to the repetition number. Ten repetitions of HRA to reference zero were undertaken with this LR movement (LR  $\Rightarrow$  0) immediately followed by ten repetitions of HRA to reference zero with a near-maximal RR movement (RR  $\Rightarrow$  0). The coordinates obtained for the different movements were plotted with different colours.

After approximately two minutes rest, a new reference zero position was established after one near-maximal amplitude LR movement of the head (with a new adjustment of the target 0,0 position). The same HRA procedure was then used to test repositioning to reference zero, in the sagittal plane, from a near-maximal Ext movement (Ext  $\Rightarrow$  0) of the head for ten repetitions and from a near-maximal Flex movement (Flex  $\Rightarrow$  0) for ten repetitions. The same sequence of movements was used for all subjects. No feedback on performance was given during the testing session. The entire procedure took approximately 30 mins.

### 2.2.5 Data Processing and Analysis

For HRA, the projection on the abscissa and ordinate axes were measured (X, Y) and each co-ordinate was given a positive or negative value according to its position relative to the corresponding axis (Figure 5). Using these two values, the subject's global HRA (R) (Heikkila and Astrom, 1996), in centimetres was then calculated trigonometrically using the following equation (2.1).

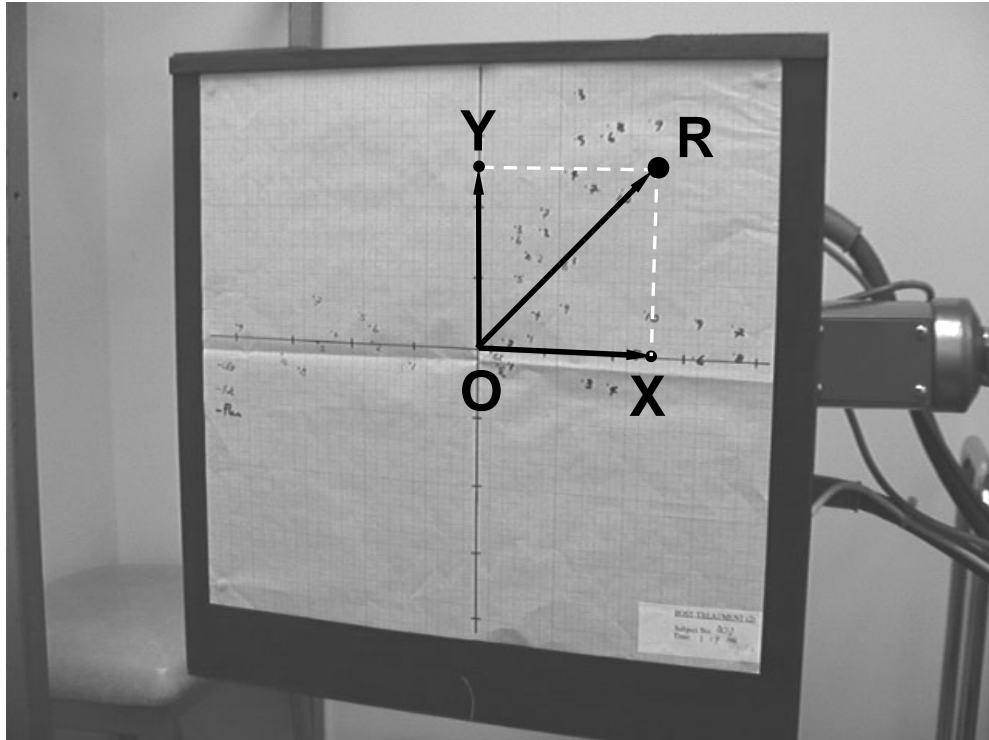


$$R^2 = X^2 + Y^2$$

$$\therefore R = \sqrt{X^2 + Y^2} \quad 2.1$$

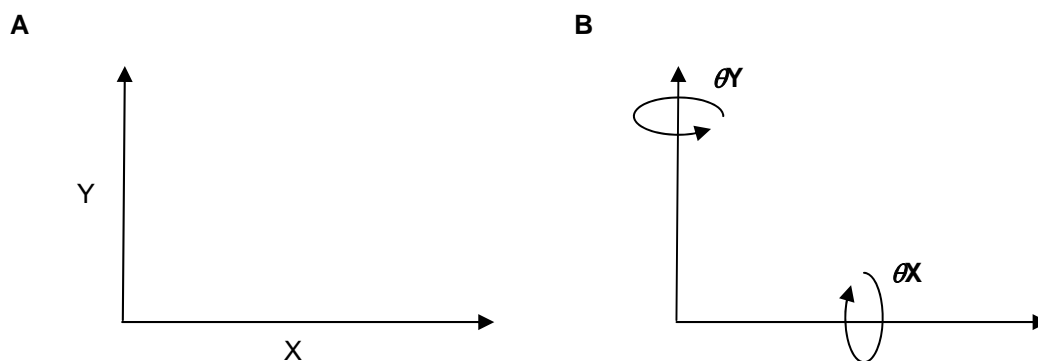
$$\text{HRA (degrees)} = \tan^{-1} \frac{\text{Repositioning Error}}{\text{Distance from target}} \quad 2.2$$

e.g.,  $\theta = \tan^{-1} \frac{X}{90}$



**Figure 5.** HRA data collection on target and initial analysis. Point O represents the centre of the target (co-ordinates 0,0), which is aligned with the projection of the light beam from the subject's reference zero position. Point R indicates the position at which the light beam stopped when the head was repositioned after a near-maximal movement. The distance O-R was converted into degrees and represents the global error of positioning (R). The horizontal projection (O-X) and the vertical projection (O-Y) indicate the horizontal and vertical components of the global error.

For every subject, mean values (centimetres) of the ten repetitions were then calculated for each HRA movement to reference zero to allow comparisons of differences between groups. Using the distance between the beam on the top of the helmet and the target, these mean centrimetric displacements of the light beam on the target were converted into angular head displacements in degrees using the following trigonometric equation (2.2). This angular data was then renamed according to the corresponding perpendicular Cartesian axes of rotation (X plane (horizontal)  $\rightarrow \theta Y$  axis of rotation and Y plane (vertical)  $\rightarrow \theta X$  axis of rotation) (Figure 6).



**Figure 6. A)** Linear (displacement) and **B)** Cartesian (angular) axes

The following commonly used measures of repositioning error were used for comparative analysis of HRA: absolute (unsigned) error (AE) and the signed error (SE). The absolute error is often presented as a logical measure of overall accuracy of performance without regard to direction. The signed error is a measure of accuracy, which takes into account the direction of error and hence is a reflection of overshoot/undershoot characteristics.

The main variables compared for differences between the patient and control groups were gender, age, active cervical ROM and HRA. A comparison of differences in gender distribution was studied using Fisher's Exact test. The interval data (age, ROM, HRA) were first examined for normality using statistical testing (Kolmogorov-Smirnov or K-S test); all the data passed this normality test ( $P > 0.1$ ). As passing the test with a relatively small sample size ( $n < 12$ ) does not necessarily mean that the values come from a Gaussian distribution a further visual inspection of the data was made using a combination of frequency histograms, Q-Q plots (Field, 2005) and box plot summaries of the measures of central tendency (mean and median) and distribution (see results section). Following this analysis, all data were considered normal in distribution and therefore parametric statistics were used for all comparisons of this interval data unless there were comparisons of more than two sets of data with unequal variances.

Differences in age between the two groups were studied using unpaired *t*-test (two-tailed). The active cervical ROM data was studied using an unpaired *t*-test (one-tailed) with Welch correction (unequal group variances). Differences in measures of HRA between the patient and control groups were studied using an unpaired *t*-test (two-tailed) with Welch correction (unequal group variances).

Secondary analyses of the differences in the global HRA and the  $\theta Y$  and  $\theta X$  components within each group, between the trials (e.g., LR vs RR vs Ext vs Flex etc.), were studied using the Kruskal-Wallis test (unequal group variances) with Dunn's post hoc testing. A further secondary analysis looking at correlations between HRA and age, intensity and duration of pain were investigated using Pearson product-moment correlation. The results have been presented using guidelines produced by Lang and Secic (2006) and Field (2005). All statistical analyses were performed using InStat<sup>®</sup> version 3.05 for Windows 95 (GraphPad Software, San Diego, CA, USA - [www.graphpad.com](http://www.graphpad.com)) and MINITAB<sup>®</sup> (Minitab Inc, State College, PA, USA). The alpha level was set at 0.05.

## **2.3 RESULTS**

### **2.3.1 Group demographics and clinical characteristics**

Over the six-week study period, fifteen new patients fulfilled the inclusion/exclusion criteria but only eleven of these patients were willing and able to participate. This cervicogenic group consisted of 6 men and 5 women, between the ages 18 to 55 yrs (mean age  $\pm$  SD = 41.1  $\pm$  13.3 yrs). Eleven healthy subjects (5 men, 6 women), range 28 to 54 yrs old (mean age  $\pm$  SD = 39.3  $\pm$  10.3 yrs) were recruited as the control subjects. A comparison of the two groups demonstrated no significant difference in gender distribution (Fisher's Exact test:  $P = 0.99$ ) and mean ages (unpaired *t*-test:  $t = 0.36$ ;  $df = 20$ ;  $P = 0.72$ ).

Data from the neck pain questionnaire profiled the cervicogenic characteristics for the patient group as follows. In all cases, the subjects described their neck pain pattern as daily or continuous. The total duration for the evolution of their pain ranged from 3 months to 5 years with a mean duration ( $\pm$  SD) of 24  $\pm$  18 months. The average intensity

of the pain (NRS) on the day of the examination procedure was  $5.1 \pm 1.9$  points (mean  $\pm$  SD) with a range of 2 to 8. Four patients (36%) reported their neck pain as predominantly left sided, 6 (55%) as bilateral/central and 1 (9%) as right sided in location.

### 2.3.2 Active range of cervical motion

An overview of the distribution of the active cervical ROM data for each group is presented in Figure 7 using box plots. The results in Table 5 show that compared with the control subjects, the patient group had a decreased active cervical range of motion (ROM) in each of the six motion directions. Unpaired *t*-tests (one-tailed) demonstrated a statistically significant difference between the two groups for RR, Flex and LLF. However, for LLF, the confidence intervals for the difference crossed zero (Table 5).

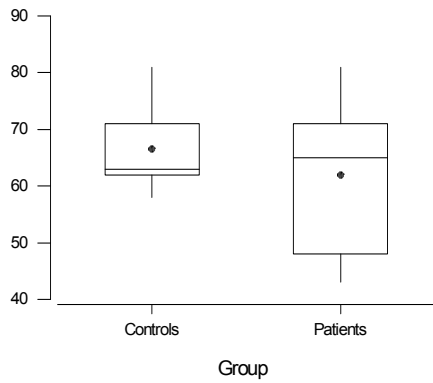
**Table 5.** Comparison of active cervical range of motion (AROM) measurements (degrees) in neck pain (n=11) and control (n=11) subjects.

	Control Subjects	Neck Pain Patients					
	Mean $\pm$ SD	Mean $\pm$ SD	d	95% CI for d	t-value	df	P-Value*
LR	66.5 $\pm$ 7.2	61.9 $\pm$ 12.8	4.6	-4.8 $\rightarrow$ 14.1	1.05	15	0.16
RR	69.5 $\pm$ 8.0	61.5 $\pm$ 7.9	8.1	1.0 $\rightarrow$ 15.2	2.38	19	0.014 <sup>†</sup>
Ext	61.6 $\pm$ 10.6	57.1 $\pm$ 13.2	4.5	-6.2 $\rightarrow$ 15.3	0.89	19	0.19
Flex	54.1 $\pm$ 8.0	44.5 $\pm$ 10.9	9.5	1.0 $\rightarrow$ 18.1	1.05	18	0.016 <sup>†</sup>
LLF	41.5 $\pm$ 5.4	33.1 $\pm$ 12.5	8.5	-0.4 $\rightarrow$ 17.3	2.06	13	0.03 <sup>†</sup>
RLF	41.2 $\pm$ 8.0	36.7 $\pm$ 12.0	4.5	-4.7 $\rightarrow$ 13.6	1.03	17	0.16

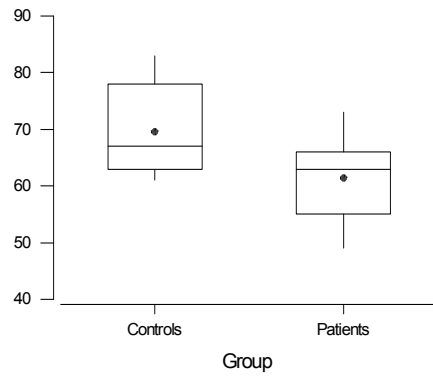
*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex), left and right lateral flexion (LLF, RLF); *d* is the mean difference (Control [mean ROM] minus Patients [mean ROM])

\**P* indicates the level of significance for comparisons using unpaired *t*-tests (one-tailed) with Welch correction. A statistically significant result is represented by  $P < 0.05$ <sup>†</sup>

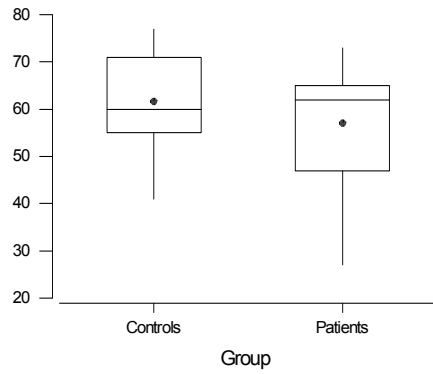
**A: Left Rotation**



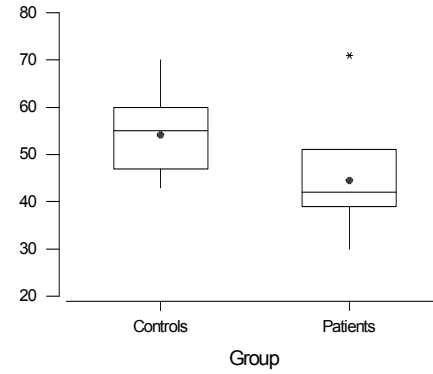
**B: Right Rotation**



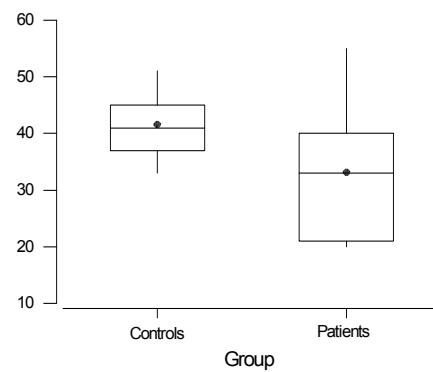
**C: Extension**



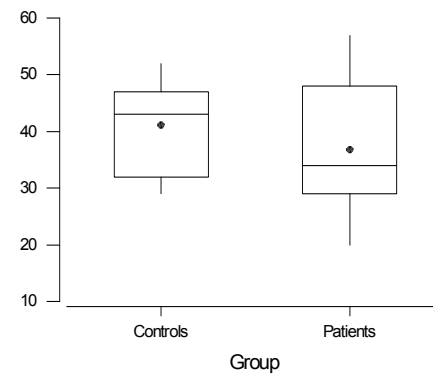
**D: Flexion**



**E: Left Lateral Flexion**



**F: Right Lateral Flexion**



**Figure 7A-F.** Box plots of active cervical range of motion data for patient and control groups in each of the movement directions. All data are presented as degrees. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \*represent outliers.

### 2.3.3 Kinaesthetic Sensibility Testing

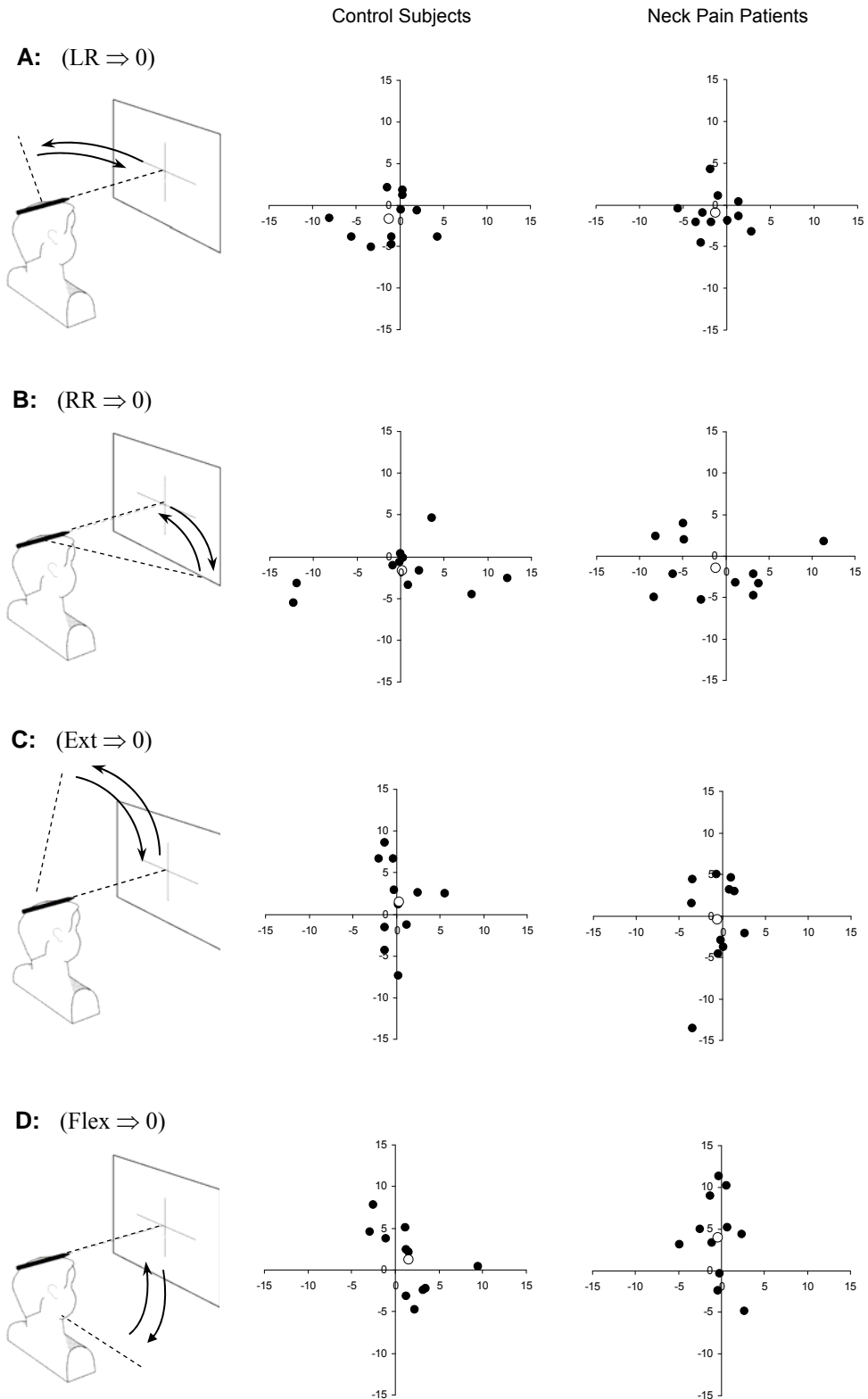
#### Primary Analyses

A graphical overview of the individual subject HRA-SSA (mean signed values of ten repositioning repetitions) for the patient and control groups, with each of the four movement directions trials, is presented in Figure 8 using scatter plots. A visual examination of the scatterplots does not demonstrate a clear general difference in cluster pattern between the groups other than a relative overshoot characteristic for neck pain group with the Flex  $\Rightarrow$  0 trial. Note with all trials except LR  $\Rightarrow$  0 for the neck pain group the largest repositioning error is apparent in the plane of primary movement.

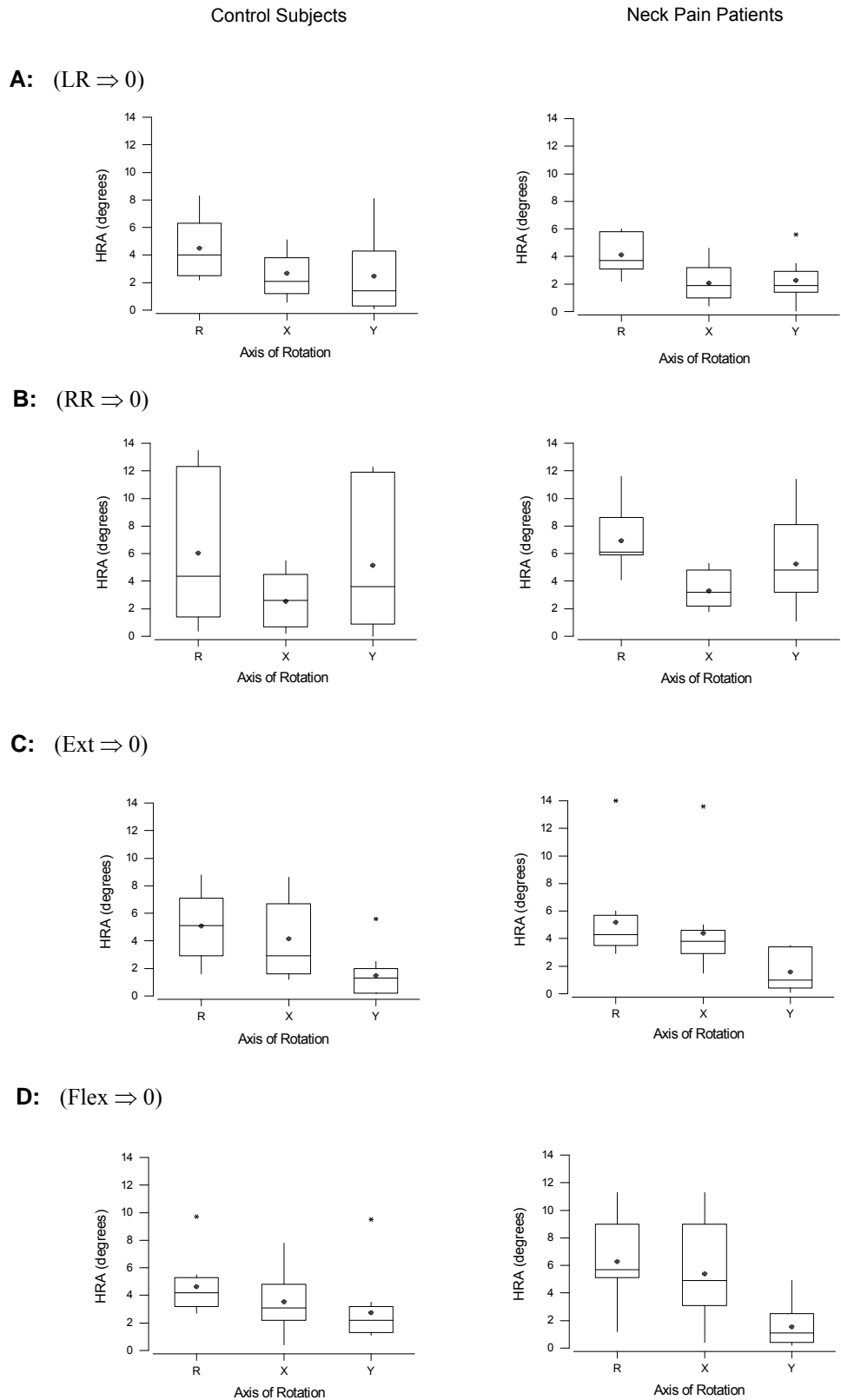
**Head repositioning accuracy (HRA): absolute values.** The mean ( $\pm$  1SD) absolute for global (R), horizontal ( $\theta$ Y) and vertical ( $\theta$ X) repositioning errors (degrees) for each group are presented in Table 6. A comparison of the distribution and measures of central tendency for the two groups is presented in Figure 9 using box plots. Comparing between groups, using unpaired  $t$  tests (two-tailed) with Welch correction (unequal group variances), there was no significant difference in the mean global HRA values (R) and the absolute mean  $\theta$ X and  $\theta$ Y components for all repositioning tests. With both groups of subjects, analysis of the absolute  $\theta$ X and  $\theta$ Y components showed that the absolute repositioning error was greater in the primary movement plane than in the perpendicular one (Table 6) for all the tests except LR  $\Rightarrow$  0. For the LR  $\Rightarrow$  0 trial this relationship was only seen in the patient group.

A secondary analysis of these within group differences, using an unpaired  $t$ -test (one-tailed) with Welch correction showed statistically significant differences between the absolute  $\theta$ X and  $\theta$ Y components for RR  $\Rightarrow$  0 ( $d = 2.0^\circ$ ; 95% CI =  $0.2^\circ$  to  $4.1^\circ$ ;  $t = 1.99$ ;  $df = 13$ ;  $P = 0.03$ ), Ext  $\Rightarrow$  0 ( $d = 2.8^\circ$ ; 95% CI =  $0.5^\circ$  to  $5.1^\circ$ ;  $t = 2.67$ ;  $df = 13$ ;  $P = 0.01$ ) and Flex  $\Rightarrow$  0 ( $d = 3.8^\circ$ ; 95% CI =  $1.4^\circ$  to  $6.2^\circ$ ;  $t = 3.43$ ;  $df = 13$ ;  $P = 0.002$ ) in the neck pain group but only Ext  $\Rightarrow$  0 ( $d = 2.7^\circ$ ; 95% CI =  $0.7^\circ$  to  $4.7^\circ$ ;  $t = 2.86$ ;  $df = 16$ ;  $P = 0.006$ ) in the control group. Within groups, comparison of the global errors, using one-way ANOVA, showed no statistically significant difference in global error between plane (vertical vs horizontal) and direction of positioning (e.g., LR vs RR).





**Figure 8.** Scatterplots showing the head repositioning accuracy (HRA) for the neck pain patients and control subjects. All data are presented as degrees. Abbreviations: HRA to reference zero following a near-maximal active head movement; left rotation (LR ⇒ 0), right rotation (RR ⇒ 0), extension (Ext ⇒ 0) and flexion (Flex ⇒ 0); The empty circle (O) on the scatterplots is the central point of the data (in plane of motion represents overshoot/undershoot characteristics in plane of movement).



**Figure 9 A-D.** Box plots of absolute global (R), horizontal ( $\theta_Y$ ) and vertical ( $\theta_X$ ) repositioning error (degrees) for the neck pain patients and control subjects. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers. For abbreviations, see Figure 8.

**Table 6.** Comparison of absolute (unsigned) Global (R), Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) in neck pain (n=11) and control (n=11) subjects

Primary Motion		Control Subjects	Neck Pain Patients	d	95% CI for d	t-value	df	P-Value*
		Mean $\pm$ SD	Mean $\pm$ SD					
LR	$\theta X$	2.7 $\pm$ 1.6	2.0 $\pm$ 1.4	-0.7	-2.0 $\rightarrow$ 0.7	0.95	19	0.36
	$\theta Y^\dagger$	2.5 $\pm$ 2.6	2.3 $\pm$ 1.5	-0.2	-2.1 $\rightarrow$ 1.7	0.20	15	0.84
	R	4.5 $\pm$ 2.1	4.1 $\pm$ 1.4	-0.4	-2.0 $\rightarrow$ 1.2	0.53	17	0.60
RR	$\theta X$	2.5 $\pm$ 1.9	3.2 $\pm$ 1.3	0.7	-0.7 $\rightarrow$ 2.2	1.10	17	0.29
	$\theta Y^\dagger$	5.2 $\pm$ 5.1	5.2 $\pm$ 3.0	0.0	-3.7 $\rightarrow$ 3.8	0.05	16	0.96
	R	6.0 $\pm$ 5.0	6.9 $\pm$ 2.2	0.9	-2.7 $\rightarrow$ 4.5	0.53	13	0.60
Ext	$\theta X^\dagger$	4.2 $\pm$ 2.7	4.4 $\pm$ 3.2	0.2	-2.4 $\rightarrow$ 2.9	0.19	19	0.85
	$\theta Y$	1.5 $\pm$ 1.6	1.6 $\pm$ 1.4	0.1	-1.2 $\rightarrow$ 1.4	0.17	19	0.86
	R	5.1 $\pm$ 2.7	5.2 $\pm$ 3.2	0.1	-2.4 $\rightarrow$ 2.6	0.10	18	0.92
Flex	$\theta X^\dagger$	3.5 $\pm$ 2.0	5.4 $\pm$ 3.4	1.9	-6.6 $\rightarrow$ 4.4	1.60	16	0.13
	$\theta Y$	2.7 $\pm$ 2.4	1.6 $\pm$ 1.4	-1.2	-3.0 $\rightarrow$ 0.6	1.41	16	0.18
	R	4.6 $\pm$ 2.0	6.3 $\pm$ 2.9	1.7	-0.5 $\rightarrow$ 3.9	1.58	17	0.13

Abbreviations: left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); d is the mean difference (Patients [mean HRA] minus Controls [mean HRA])

$^\dagger$  Primary axis of rotation for the repositioning test

\*P indicates the level of significance for comparisons between groups using unpaired t-tests (two-tailed) with Welch correction. A statistically significant result is represented by  $P < 0.05$

**Head repositioning accuracy (HRA): Signed values.** Table 7 presents the signed data used for over/undershoot error analysis and the distribution and measures of central tendency for the two groups is presented in Figure 10 using box plots. This index of HRA is also represented graphically by the empty circle on the scatterplots ( O ) shown in Figure 8. One-sample t-tests (two-tailed) comparing the means to reference zero, showed no significant over/undershoot tendencies in the controls for any of the repositioning directions. Within the patient group, a significant overshoot (+ve  $\theta X$  [vertical plane]) was found for Flex  $\Rightarrow$  0 ( $d = 4.0^\circ$ ; 95% CI =  $0.6^\circ$  to  $7.4^\circ$ ;  $t = 2.59$ ;  $df = 10$ ;  $P = 0.027$ ). Unpaired t-tests (two-tailed), with Welch correction, demonstrated no significant differences between the groups (Table 7).

**Table 7.** Comparison of signed Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) neck pain (n=11) and control (n=11) subjects<sup>†</sup>

Primary Motion		Control Subjects	Neck Pain Patients	d	95% CI for d	t-value	df	P-Value*
		Mean $\pm$ SD	Mean $\pm$ SD					
LR	$\theta X$	-1.7 $\pm$ 2.7	-1.0 $\pm$ 2.3	0.7	-1.5 $\rightarrow$ 3.0	0.68	19	0.51
	$\theta Y^{\ddagger}$	-1.2 $\pm$ 3.4	-1.3 $\pm$ 2.5	-0.1	-2.7 $\rightarrow$ 2.6	0.04	18	0.97
RR	$\theta X$	-1.6 $\pm$ 2.8	-1.5 $\pm$ 3.4	0.1	-2.6 $\rightarrow$ 2.9	0.14	19	0.89
	$\theta Y^{\ddagger}$	-0.2 $\pm$ 7.3	-1.1 $\pm$ 6.2	-0.9	-7.0 $\rightarrow$ 5.1	0.32	19	0.75
Ext	$\theta X^{\ddagger}$	1.5 $\pm$ 4.9	-0.5 $\pm$ 5.6	-2.0	-6.7 $\rightarrow$ 2.7	0.89	19	0.39
	$\theta Y$	0.3 $\pm$ 2.2	-0.5 $\pm$ 2.1	-0.8	-2.7 $\rightarrow$ 1.1	0.88	19	0.39
Flex	$\theta X^{\ddagger}$	1.2 $\pm$ 4.0	4.0 $\pm$ 5.1	2.7	-1.4 $\rightarrow$ 6.8	1.40	18	0.17
	$\theta Y$	1.6 $\pm$ 3.4	-0.4 $\pm$ 2.1	-2.0	-4.5 $\rightarrow$ 0.6	1.64	16	0.12

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); *d* is the mean difference (Patients [mean HRA] minus Controls [mean HRA])

<sup>†</sup> Values in the primary axis of rotation represent overshoot and undershoot characteristics relative to the target's centre (reference 0) and direction of repositioning movement.

<sup>‡</sup> Primary axis of rotation for the repositioning test

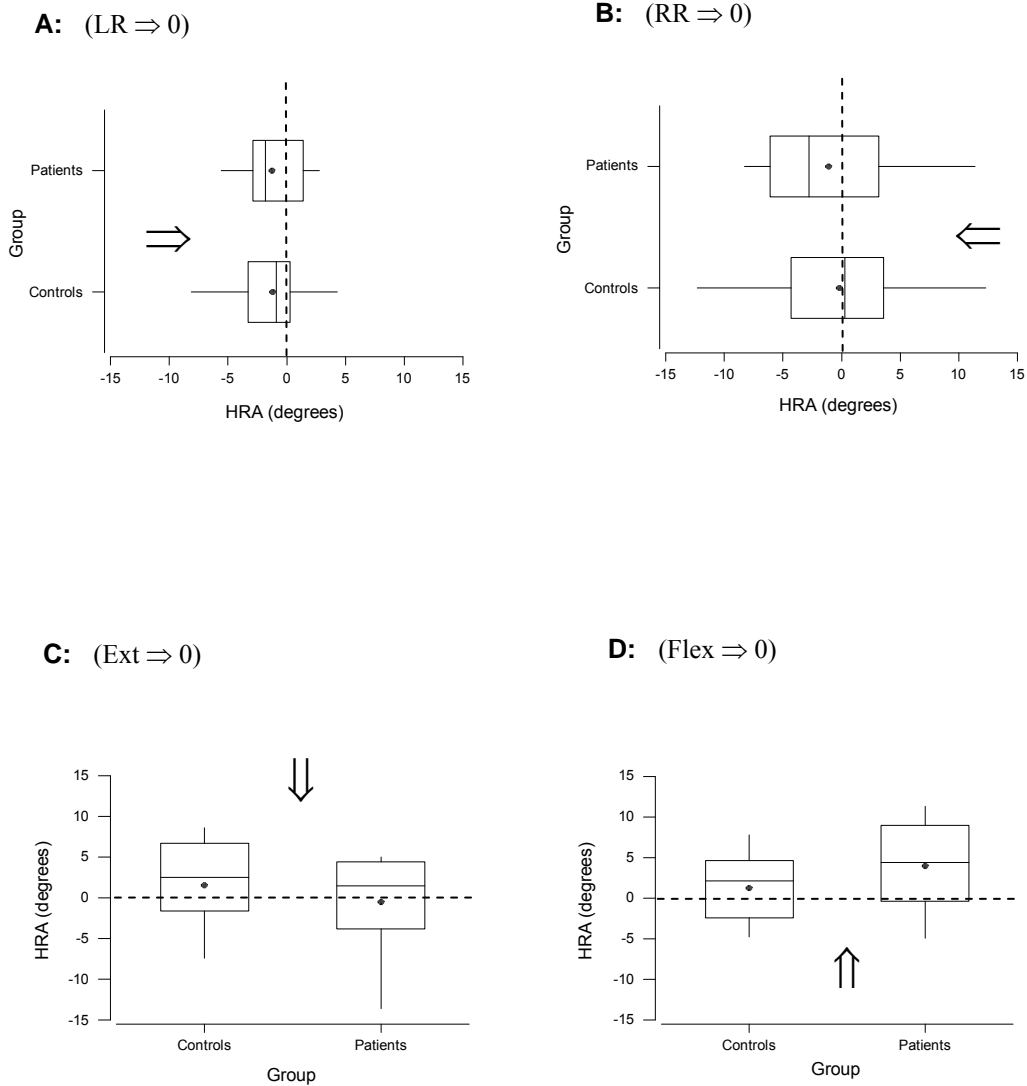
\**P* indicates the level of significance for comparisons between groups using unpaired *t*-tests (two-tailed) with Welch correction. A statistically significant result is represented by *P* < 0.05

## Secondary Analyses

**Variability of HRA between trials.** Analysis of the mean global HRA values (R) and the absolute values for the  $\theta Y$  (horizontal) and  $\theta X$  (vertical) components (Table 6), using the Kruskal-Wallis test (unequal group variances) with Dunn's post hoc testing, demonstrated that there were no significant differences within the control subject group between the four trials in the different planes. Analysis of the patient group values, demonstrated a significant difference between the mean global HRA values (R) for RR  $\Rightarrow$  0 vs LR  $\Rightarrow$  0 ( $p < 0.01$ ) and RR  $\Rightarrow$  0 vs Ext  $\Rightarrow$  0 ( $p < 0.05$ ). However, this pattern of HRA variability was not suggestive of learning interference, fatigue, and drop in the attention span or other such phenomena. The same statistical analysis of the signed data also showed no significant differences in overshoot/undershoot tendencies between the four trials.

***Variability of HRA within trials.*** Throughout the series of repetitions for each of the repositioning movements, no trends or statistically significant differences (using the Kruskal-Wallis test [unequal group variances] with Dunn's post hoc testing) suggestive of learning interference, fatigue, drop in the attention span or other such phenomena were seen in the absolute HRA for both groups. Interestingly, for both groups of subjects, any error of head repositioning essentially occurred with the first of the ten repetitions. No significant deviation from this relocation point was found with the remaining nine series of repositionings (using the Kruskal-Wallis test [unequal group variances] with Dunn's post hoc testing). A similar pattern was seen with the signed data with no significant differences (drift or fatigue etc.) in over/undershoot tendencies throughout each series of repetitions

***Correlation of clinical variables with HRA.*** There was no evidence of correlation between head repositioning error and age for either of the study groups nor was there any correlation between head repositioning error in the neck pain group and pain characteristics (pain intensity on day of examination and duration of pain).



**Figure 10.** Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients and control subjects. The horizontal ( $\theta_Y$ ) and vertical ( $\theta_X$ ) repositioning error (degrees) are displayed for L & RR  $\Rightarrow$  0 (A-B) and Ext & Flex  $\Rightarrow$  0 (C-D) respectively. For L & RR, the middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25th and 75th percentile); For Ext & Flex, the middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot.; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005).

### 2.3.4 Summary of Kinaesthetic Sensibility Primary Results.

#### Head repositioning accuracy to subject straight ahead (HRA-SSA)

- **Absolute HRA:** No significant difference between control and neck pain group
- **Signed HRA (overshoot/undershoot):** A significant overshoot with the Flex  $\Rightarrow$  0 trial for the neck pain group. No significant difference between the control and neck pain groups.

## 2.4 DISCUSSION

This preliminary study focused on cervicocephalic kinaesthetic sensibility in one subgroup of neck pain patients; namely, those with chronic neck pain of non-traumatic (insidious) onset. More specifically, the ability of the subjects to actively return the head to a subjective 'straight-ahead' position was examined (reference zero – obtained after a movement of the head in the perpendicular plane to the primary plane of movement for the trial). The overall results from the limited sample did not indicate a general reduction in HRA within the neck pain group compared with the healthy age and sex matched controls. These results are consistent with Teng et al., (2007) but seem to conflict with the results of Kristjansson et al., (2003) and Sjolander et al., (2008). As with many comparisons of cervicocephalic kinaesthetic sensibility studies, variations in results may reflect different clinical characteristics in the neck pain populations investigated and differences in testing protocols and measuring equipment.

### **Patient Group and HRA**

With respect to patient group characteristics, the patients in this study were most comparable with those investigated by Kristjansson et al., (2003) in terms of pain duration. However, the sample size was smaller in the current study and the mean age ( $\pm$  SD) of the patients was older ( $41.1 \pm 13.3$  years vs  $30.0 \pm 13.3$  years). The patients studied by Sjolander et al., (2008) were of a similar mean age as the current study but with a much longer mean pain duration ( $97 \pm 68$  months vs  $24 \pm 18$  months). It is also interesting to note that the pain severity (mean  $\pm$  SD) may have been higher in the neck pain patients with current study compared with Kristjansson et al., (2003) ( $5.1 \pm 1.9$  points vs  $3.15 \pm 2.1$ ). However, different pain rating scales (NRS vs VAS and characteristics ('last few days' vs 'past week')) were used in the studies. In asymptomatic subjects, a worse HRA to NHP has been demonstrated with the sagittal plane trials (Flex/Ext) for 'middle-aged' subjects compared with young adults (Teng et al., 2007). A reduced HRA to NHP has also been reported in 'whiplash' patients 3 months post-injury when experiencing moderate/severe neck pain and disability (NDI > 30) compared with those experiencing mild neck pain and disability (NDI 10-28), recovered subjects (NDI <8) and asymptomatic controls (Sterling et al., 2003). However of the three directions tested (RR  $\Rightarrow$  0, LR  $\Rightarrow$  0, Ext  $\Rightarrow$  0), statistical significance was demonstrated for the



right rotation trial only between the moderate/severe subjects and the other groups (Table 1). The effect of age and perceived pain severity has yet to be solidly established particularly with insidious onset neck pain.

### **Measurement Protocols & HRA**

The conflicting results could be explained by differences in the measuring methods/protocols and equipment. As highlighted in the introduction, all studies that have investigated the cervicocephalic kinaesthetic performance in patients specifically with insidious onset neck pain, have utilised a repositioning task to the central position (NHP) (Kristjansson et al., 2003; Palmgren et al., 2006; Teng et al., 2007; Sjolander et al., 2008). However there are some obvious and more subtle differences in testing protocols. The method used by Kristjansson et al., (2003) closely resembled the original method of testing introduced by Revel et al., (1991) but only 3 trial repetitions were used to arrive at the mean repositioning accuracy scores (Table 1 – Chapter 1).

Sjolander et al., (2008) tested the subjects whilst standing and used 8 repetitions. The method used in our study differed in several respects. The repositioning target used for this study was defined as a subjective straight ahead position of the head, a position found after a full active movement of the head in a perpendicular plane. Kristjansson et al., (2003) and Sjolander et al., (2008) utilised the neutral head posture as target but this was not determined after a specific head movement; the subjects, as in most HRA-NHP studies were simply asked to assume a NHP and remember that as the target.

The methods may also have differed in that the subjects in the current study were asked to perform a near-maximal movement of the head for the repositioning task instead of a “full active rotation of the head and neck within comfortable limits” (Kristjansson et al., 2003) or to “comfortably” rotate the head as fast and as far as possible” (Sjolander et al., 2008). The rationale for the method in the current study was that most of the neck pain subjects experienced a sharp increase in pain at the end range of motion which it was felt could possibly bias the subjects repositioning ability compared with the controls. It was also apparent during the pilot trials, that when asked to turn the head maximally, a degree of shoulder and trunk rotation took place. It was felt that this needed to be removed in order to maintain neck isolation.

A consistent initial perpendicular plane head movement was utilised when finding the SSA position for each of the horizontal and vertical movement trials. A limited examination of the data did not indicate that there was any interaction of head repositioning error between the movement planes but further investigation of the effect of the preliminary movement on the HRA may be warranted.

Lastly, ten repetitions were utilised to obtain a mean value in the current study and a new reference position for SSA was found only when moving from the horizontal to the vertical plane tests rather than after each task. Overall it is not clear if these variations could affect the validity or affect the ability of the tests to detect abnormalities. Certainly many variations of SSA or NHP tests have been found to be discriminative for HRA between other patient populations and control groups (Table 1). The significance of these methodological differences may need to be followed up in future studies.

### **Head Movement Velocity and HRA**

It may be important to consider some other possible issues of testing methodology that may affect the discriminative value of the test in the current study. The cervical kinaesthetic test used in this and other studies is thought to examine primarily cervical spine kinaesthetic performance although a recent study has cast doubt on this specificity (Demaille-Wlodyka, et al., 2007). As highlighted in Chapter 1, when considering movement of the head relative to a stationary trunk ('head-on-trunk' movements), the integrated nature and functional overlap of the various sensory inputs within the systems involved in equilibrium, spatial and self-motion awareness makes isolation of one sub-system, such as the cervical mechanoreceptive apparatus, difficult. Results of studies in which subjects graded their perception of various movements, suggest that the head movements undertaken in this and previous studies, particularly in the horizontal plane may utilise peripheral kinaesthetic information primarily from the cervical spine mechanoreceptive apparatus (Mergner, Nardi, Becker and Deecke, 1983). In addition, there is further evidence to suggest that this may only be the case in low speed head movements (Mergner et al., 1983; Mergner, Siebold, Schweigart and Becker, 1991; Mergner, Hlavacka and Schweigart, 1993). As no speed instructions were given during our testing procedure, the possibility exists that no clinically meaningful differences were found because the active movements in both groups conducted too rapidly

effectively, rendering the test insensitive to any impairment in cervical proprioception. However, the subjects studied by Kristjansson et al., (2003) were also not given speed instructions and those investigated by Sjolander et al., (2008) were asked to move as fast as possible.

To the authors knowledge only one recent study testing HRA-NHP has carefully controlled for head movement speed and this failed to find a difference in HRA (Teng et al., 2007). It is possible that the rate of movement may not be a methodologic issue for repositioning tasks that involve a cessation of movement at the relocation target such as HRA-SSA as most of the tonic neuronal component (corresponding to head position) in the test procedure may be identified as cervical (Taylor and McCloskey, 1988). At this stage in our understanding of the neurophysiological mechanisms involved in the various kinaesthetic tasks and as other investigators have acknowledged, further research is needed to determine the types of kinaesthetic tasks and speeds of movement that most specifically challenge the neck mechanoreceptors and related pathways more than receptors in the vestibular apparatus (Kristjansson et al., 2004).

### **Cervical Spine Range of Motion and HRA**

The effect of cervical active range of motion (AROM) during the testing may also be a consideration when performing cervicocephalic kinaesthetic tasks. The patients in the current study exhibited a slightly reduced AROM in most directions. However it was not measured simultaneously with the HRA testing. Although the subjects were asked to perform a near-maximal movement of the head there is a possibility that a reduced range of movement in the test was related to increased accuracy. As such, differences between a patient group with a reduced AROM and a control group could be diminished when the repositioning test requires a full head movement. Sjolander et al., (2008) did find ROM during the HRA testing to be a significant covariate for the errors found. It has been argued that using an end range of motion as the starting position for the relocation movement may bring a confounding factor into the test procedure if there are systematic changes in end range that occur during testing (Lee, Nicholson, Adams and Bae, 2005). The authors suggest that this may occur in neck pain patients through a sensitization effect and result in a reduced range for the next repetition and lead to overshooting the target. They also suggest that this may be an explanation for the “overshooting”

tendency that has been reported in studies utilizing repositioning task from an end or near-end range of motion position (Revel et al., 1991; Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998; Treleaven et al., 2003). No significant differences in overshoot/undershoot were found between the subject groups in the current study suggesting that a movement sensitisation variable may not have been a factor in this case. As yet, the relationship between AROM during testing and HRA performance has yet to be directly defined.

### **Control Subjects and HRA**

Another consideration when examining the lack of difference between patient and control subjects in the present study is a poor HRA performance from the control subjects i.e., they don't reflect a true normal population. Care was taken to screen the subjects carefully for confounding variables and there were no clear outlying performances from this group which could significantly affect the small sample size in this study (Figures 8-10). Subject motivation may also be a factor in HRA performance. It is possible that the control subjects were less focused and motivated than the patient group and perhaps concentrated less thus relatively underperforming in the HRA task. Beyond this it is hard to judge the control subject performance without comparison to other studies. This is fraught with difficulties due to differences in methodologies and measurement equipment. It is also hard to compare in detail as many studies have not provided all the measurement parameters used in this study or presented them in different units (Table 1 – Chapter 1).

### **Measurement Instrumentation and HRA**

The laser pointer method of measuring cervicocephalic kinaesthetic sensibility used in this and many of the other HRA-NHP studies (Table 1 – Chapter 1) utilises a relatively simple equipment design which was inexpensive, easy to execute and may permit a degree of discriminant classification of certain neck pain subgroups. However, the method of measurement and in particular, its subjective and non-remote nature inevitably involves a degree of experimenter bias and geometric variations and inaccuracy. Indeed as Humphreys and Irgens (2002) point out, marking points on a flat target may not be the most accurate method of measuring HRA as the head moves in an arc. Although on the surface there appears to be a reasonable reproducibility and

consistency within study settings and similar neck pain populations, formal investigation and adequate statistical reporting of the HRA-SSA testing characteristics using this method, such as test-retest *reliability*, has been limited (Revel et al., 1991; Revel et al., 1994; Heikkila and Astrom, 1996). With this cautionary note in mind, it is interesting to observe that with a few examples, the control subjects appeared to exhibit a much poorer global HRA in the current study than originally reported by Revel et al., (1991) and subsequent studies using a laser pointer method (Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998; Humphreys and Irgens, 2002; Kristjansson et al., 2003; Treleaven et al., 2006). This suggests a poor performance of the control subjects in the present study.

It is also interesting to note that the neck pain group global error was also greater than the error reported in two of the studies (Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998) despite these studies finding impaired HRA in the patient group, comparable with that reported by Revel et al., (1991) but smaller than that found with Humphreys and Irgens (2002). If the original overall global HRA is considered, based on the radial 4.5° threshold value (11% chance of healthy subjects outside this zone) between neck pain and healthy groups described by Revel et al., (1991) in the original study of HRA-NHP, 48% of the healthy subject values (21 of 44 measurements) in the present investigation were outside this zone. Using the same radial 4.5° threshold value, 38% (17 of 44 measurements) of the patient global HRA values were inside this zone. These comparisons serve to further highlight the difficulties in comparing HRA values particularly between different study settings. They also reinforce the need for more research into the fundamental clinical characteristics of the HRA-SSA test and a better understanding of normative performance.

### **Sample Size and Statistical Error**

Before concluding that the patient group in the present study did not have a deficit in HRA-SSA associated with their condition, the possibility of a Type II statistical error (false –ve) also needs to be considered. It may be that the lack of kinaesthetic deficit compared with previous investigations is not wholly attributable to differences in the neck pain causation but the result of a relatively small sample size and/or a large scatter of the data with insufficient statistical power to reveal any differences.

The sample size in the present study ( $n = 11$ ) created difficulties in deciding whether to use parametric or non-parametric statistical tests. One of the first criteria for the use of many parametric statistics is that the data is normally distributed. Observation of the data samples using the box plots shown in the results section do show several data distributions that are skewed. Passing the normality testing (Kolmogorov-Smirnov test) with small sample sizes does not necessarily mean that the values come from a Gaussian distribution. Small sample sizes may not provide sufficient data to allow a reasonable discrimination between Gaussian and nongaussian distributions. If the population is far from Gaussian, the  $P$ -value may be misleading with parametric tests (Type I error – false +ve). It could therefore be argued that these issues coupled with the unequal variances of some of the data sets being compared are reasonable grounds for using the more ‘conservative’ nonparametric statistical analysis throughout. However, although nonparametric tests may be appealing because they make fewer assumptions about the distribution of the data, they are less powerful than parametric tests. They usually result in higher  $P$ -values and may make it harder to detect real differences particularly when the sample size is small (Type II error).

The decision to use parametric statistics for the majority of the analyses in this study was a balanced one based on the normality testing coupled with observation of the data distributions and measures of central tendency (means vs medians) using box plots, frequency histograms and Q-Q plots. Although no significant differences were found using this relatively ‘liberal’ approach, a secondary descriptive analysis of the measures of central tendency did reveal lower median overall global HRA values for both the patient and controls groups (Patients -  $5.0^\circ$ ; Controls –  $4.8^\circ$ ) compared with the mean values used for analysis here (Patients –  $5.7^\circ$ ; Controls –  $5.1^\circ$ ). When analysed using non-parametric statistics (Mann-Whitney  $U$  test), a statistically significant reduced global HRA (R) was seen in the neck pain group for Flex  $\Rightarrow$  0 compared with the control subjects. However, the fact that no significant difference could be found with the Ext  $\Rightarrow$  0 and horizontal rotation trials (particularly in the plane of motion) and the lack of distinct differences in the clustering of the data displayed in the scatterplots would suggest this statistical difference has limited clinical meaning.

### **Order of Testing and HRA**

Some other points from this study are perhaps worth noting. Some HRA studies have utilised the same order of testing directions for each subject (Revel et al., 1991; Sterling et al., 2003; Treleaven et al., 2003) although in many, this aspect of the testing protocol is not clearly stated in the methodology (Revel et al., 1994; Heikkila and Astrom, 1996; Rogers, 1997; Heikkila and Wenngren, 1998; Heikkila et al., 2000; Soderlund et al., 2000; Kristjansson et al., 2001; Humphreys and Irgens, 2002). Revel et al., (1991), reported a relatively poor HRA-SSA performance for the healthy subjects manifesting as a relatively high percentage of subjects overshooting the reference zero point for Flex  $\Rightarrow$  0 compared with other directions. As this was the last of the four trials, the authors felt that it may be due to a drop in attention span as the trial sessions had already lasted approximately 30 minutes at the point of testing Flex  $\Rightarrow$  0. A similar trend was seen for the mean Flex  $\Rightarrow$  0 signed data with the present study as stated earlier no significant statistical difference was found (Table 7 and Figures 8 & 10), randomising the order of trials in future studies may be warranted to help minimise any possible effects of order on HRA-SSA performance as has been adopted by some investigators (Kristjansson et al., 2003; Treleaven et al., 2006).

### **Head Position Endpoint and HRA**

An interesting observation was also noted in the current study when subjects relocated their head position to the target with each trial repetition. A short-lived wobble or fine drift was seen immediately on reaching the repositioning target followed by a more stable control of the head position. As such, the coordinates were recorded after a short latency (approx 2 secs) to allow this drift to settle and reduce errors potentially caused by this movement. This movement phenomenon has also been reported by Feipel et al., (2006) and Palmgren (2006) the later of which incorporated a two second latency with his recording method. It might be interesting to study this observation using the more sophisticated 3-D measuring devices which might allow a more accurate assessment of the subject's true position of relocation (Feipel et al., 2006). Some researchers have asked subjects to indicate when they have reached the NHP before recording the coordinates (Treleaven et al., 2006; De Hertogh, et al., 2008).

### **Initial Head Position and HRA**

A final observation from the neck pain subjects in the current study involved the initial position assumed as the reference zero for the repositioning task. Several of the neck pain group were clearly not in as 'straight-ahead' position with respect to the trunk. Many appeared to be in a relatively extended or rotated neck position. This has been noted by Loudon et al., (1997) when looking at repositioning to non-neutral set points within the range. It is possible that neck pain subjects are poor at finding a SSA position but relatively better at relocating to this position (what they think is SSA). As such, assessing actual head position against true 'straight-ahead' position might yield some interesting discriminative information between patients and normal subjects.



## 2.5 CONCLUSION

The results of this study suggest that patients with chronic neck pain of insidious onset (non-traumatic) show little evidence of impaired cervicocephalic kinaesthetic sensibility, when measured as HRA to a subjective 'straight-ahead' position. This is contradictory to subsequent investigations of this neck pain subgroup (Kristjansson et al., 2003; Sjolander et al., 2008). The conflicting results may reflect a genuine difference in the patient populations highlighting further the possible heterogeneity of chronic 'mechanical' neck pain. However this conclusion may be premature and consideration for the somewhat contradictory results should also be given to differences in methodology and specific aspects of measurement protocol. Further investigation of cervicocephalic kinaesthetic sensibility performance in patients with chronic insidious neck pain is certainly warranted possibly with further comparisons with 'whiplash' patients (Kristjansson et al., 2003; Sjolander et al., 2008). There is also a need for investigation into the effects of differing measurement protocols on kinaesthetic performance and the discriminative value of tests of HRA-SSA/NHP. This study also highlighted the need for establishing a more definitive normative knowledge base, testing characteristics and the difficulties in comparing values between settings with the current non-remote measuring procedure.

## **Chapter 3**

---

# **HEAD REPOSITIONING ACCURACY: A CRITICAL REVIEW OF 'RELIABILITY' CHARACTERISTICS AND METHODS OF STATISTICAL ANALYSIS**

---

### **3.1 INTRODUCTION**

The ultimate criterion for the usefulness of a diagnostic test is whether it adds information beyond that otherwise available and does this information lead to a change in clinical management that is ultimately beneficial to the patient (Jaeschke, Guyatt and Sackett, 1994). As already highlighted, prior to the start of this course of study there was some evidence to suggest that measurement of head repositioning accuracy to subjective straight ahead (HRA-SSA) may be a useful tool for the analysis of cervicocephalic kinaesthetic sensibility deficits. (Revel et al., 1991; Revel et al., 1994; Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998).

The presence of abnormalities in cervicocephalic kinaesthetic sensibility may be one functional factor responsible for the persistence or recurrence of neck pain symptoms in some patients. However, the study reported in Chapter 2 found little difference between the control and neck pain groups. This raised the possibilities that there may be no clinically meaningful abnormality in HRA-SSA in this particular subgroup of patients with mechanical neck pain, or that the current test of kinaesthetic sensibility is not particularly discriminative in this population. The study also served to highlight, several difficulties with the current method of measurement and the relatively poor knowledge of HRA in healthy subjects. Many clinical tests are adopted prematurely because they have not been adequately evaluated. With tests of HRA-SSA, there is a need for normative data and adequate evaluation of the clinical characteristics of the measurement parameter and measuring methods.

## 3.2 THE DEVELOPMENT AND CHARACTERISTICS OF CLINICAL TESTS

There are four broad guideline steps that where appropriate should be followed when developing, describing and promoting the adoption of a new clinical test (Table 8).

### Purpose of the Test

Identifying the purpose of the test is fundamental. A clinical test is used to perform a specific *function* in a specific *population* that is believed to have a specific *condition or abnormality*. It is important to describe each of these components. Tests usually have one or more functions (Table 9). In the current context, measurements of cervicocephalic kinaesthetic parameters have been proposed as diagnostic and monitoring tools.

**Table 8.** Guidelines for developing, describing and promoting the adoption of a new clinical test.

---

#### **Purpose** of the test

- Identify the purpose of the test (see Table 9).
- Specify the stage of the condition (disease) for which the test is appropriate.
- Explain the meaning of a positive test result.

#### **Characteristics** of the test

- Describe the biological principle on which the test is based.
- Report the validity of the index test that is under study and the reference test to which it was validated.
- Report reliability of the test.
- Explain the meaning of equivocal results and how such results were incorporated into the calculation of the test's characteristics.
- Report the diagnostic sensitivity and specificity of the test, including the associated (95%) confidence intervals.
- Report the positive and negative likelihood ratios of the test.

#### Clinical **Application** of the test

- Describe how the test is to be administered.
- Report the positive and negative predictive values of the test, as well as the prevalence of the disease associated with these values.

#### Considerations in **Adopting** the test

- Describe the human, financial, and physical resources necessary to offer the test in a given setting.
- Describe the medical costs and benefits of adopting the test.
- Describe the financial costs and benefits of adopting the test.
- Describe how the test compares with similar tests.

---

From Lang and Secic (1997)

It must be emphasised that abnormalities in cervicocephalic kinaesthetic sensibility do not in themselves represent a marker for a clinical diagnosis such as chronic ‘mechanical’ neck pain. The test is really aimed at testing for functional abnormalities that may be a contributing pathological factor in some patients with mechanical neck pain or other mechanical cervical spine related syndromes. It could be argued that in the case of the more controversial cervical somatosensory dizziness, an abnormality in cervicocephalic kinaesthetic sensibility is a direct diagnostic indicator if it identifies aberrant cervical spine mechanoreceptive function and this is not present in other dizziness pathologies.

**Table 9.** Functions of clinical tests

A <b>Screening</b> test	Performed on healthy asymptomatic people; used to identify those who are at risk of a specific disorder; outcome may justify a subsequent diagnostic test or direct preventative action; a good screening test has high <i>sensitivity</i> .
A <b>Routine</b> test	Performed on symptomatic subjects; used as part of a battery of tests and may result in a ‘finding’ that is unrelated to the presenting condition.
A <b>Diagnostic</b> test	Performed on symptomatic subjects; used specifically either to identify the presence or absence of a disorder: a good diagnostic test has high <i>specificity</i> .
A <b>Staging</b> test	Performed to quantify and characterise the nature or extent of a condition.
A <b>Monitoring</b> test	Performed to track the progress of a condition over time.

From Lang and Secic (1997)

The specific population for which the test is appropriate should also be identified with indications of the necessary demographic, diagnostic and comorbidity characteristics. These are often not explicitly defined when tests are adopted. Tests are often developed using studies with relatively narrow inclusion/exclusion criteria but then utilised in broader populations without further development studies. The diagnostic characteristics of the population under question are intimately wrapped up with the clinical condition under consideration. To date, HRA tests have been applied in adults with ‘mechanical’ neck pain and also the more controversial cervical somatosensory dizziness. It should be realised that some tests may differ in their ability to detect early and late forms of disease or functional abnormality. The tests may also perform poorly when used in a population with the full spectrum of a disease but may be more discriminative in certain subgroups where the disorder is more narrowly defined. For instance, the HRA test may

not be very discriminative in a population of patients defined only as having ‘mechanical’ neck pain. This group potentially includes all stages of the condition (acute to chronic), all severities (pain or other parameter) and all aetiologies (insidious vs trauma related). If subgroups have different pathological components and the population under investigation is too heterogeneous, then any abnormalities might be diluted or hidden. In addition to these limitations, the test may also perform well as a diagnostic or screening test but may not exhibit the properties to measure smaller changes when monitoring a patient’s progress over time.

### **Characteristics of the Test**

Establishing the characteristics of a test is a vital step (Table 8). These characteristics help determine its usefulness and which functions the test is best suited to (Table 9). It should be highlighted that the problem at this stage of development for cervicocephalic kinaesthetic sensibility testing is that we are still investigating the existence and nature of any abnormality. It is difficult to develop and describe the properties of a test if you don’t know whether all the subjects have the abnormality or not. An important first step in establishing the usefulness and efficacy of any assessment tool is the investigation of its *reliability* (Haas, 1991a; Maffey-Ward et al., 1996; Atkinson and Nevill, 1998; Lachin, 2004).

In simple terms, *reliability* refers to the level of agreement or an estimate of the measurement error and bias between measurements. It has been argued that *reliability* should be quantified before *validity* since it will never be valid if the measurements are not adequately reproducible or repeatable (Atkinson and Nevill, 1998). However, it should be pointed out that acceptable *reliability* does not confer *validity*. *Reliability* is fundamental to all aspects of clinical research. Without it there cannot be confidence in the data collected nor can rational conclusions be drawn from those data (Ottenbacher, 1995). The uses of *reliability* estimates include 1) decision making when monitoring individuals 2) comparison of tests or equipment 3) estimation of sample size in experiments 4) the eligibility criteria and choice of a primary outcome measure for a clinical trial and 5) estimation of the magnitude of individual differences needed in response to treatment (Hopkins, 2000; Lachin, 2004).

The term *reliability* is often applied to several types of what could be seen as agreement investigations. Under this *reliability* umbrella, three main study types are of particular interest to clinicians and appear most frequently in the movement science and manual therapy literature (Table 10). These study types will form the basis of the remaining discussion.

Study designs 1 and 2 are commonly referred to as *test-retest* (or *intra-observer/rater*) and *inter-observer/rater reliability* respectively. Studies comparing different methods (study design 3) are increasingly referred to as method agreement studies. When studying *reliability*, it must be appreciated that there will always be some error and variability in measurement particularly where biological variables are the subject of investigation. Therefore, it is very important that *reliability* should be viewed as the amount of measurement error that is acceptable for the effective specified use of the measurement (Bland and Altman, 1986; Atkinson and Nevill, 1998). How variable the measurements can be without causing difficulties is a matter of judgement. It will depend on the purpose of the study, the variable being measured, the resolution of the instruments and the differences that would be considered clinically meaningful in a trial or experiment (Bland and Altman, 1986). This concept is important when considering the most appropriate method of *reliability* assessment and the interpretation of results.

**Table 10.** Terms used to describe common *reliability* study designs Devised by Webb and Rix (2005), used with permission.

<i>Reliability</i> Study design	Terms	
	Prefix	Suffix
1) Agreement between measurements made by <u>same method</u> and <u>observer</u>	Intra-session/examiner/rater/tester/observer Test-Retest	concordance, consistency, reliability, agreement, precision, accuracy, stability, reproducibility, repeatability
2) Agreement between measurements made by <u>same method</u> and <u>different observer</u>	Inter-examiner/rater/tester/observer	
3) Agreement between measurements made by <u>different methods</u> (same observer)	Method agreement, Instrument reliability, concurrent validity	

*Reliability* studies are numerous in the literature exhibiting a diversity of experimental designs, terminologies and statistical approaches (Atkinson and Nevill, 1998; Rankin and Stokes, 1998; Bruton, Conway and Holgate, 2000). However, there is often inconsistency with this terminology and many examples of inappropriate statistical analyses and inferences particularly with studies examining interval data. The issues surrounding terminology are perhaps not as meaningful as the statistical analyses and inferences, but nevertheless are still a source of confusion. The prefix terms are generally used to describe the source of error or variability under investigation (Table 10). However, these terms are often applied imprecisely, used interchangeably and don't always specifically reflect the potential sources of error/variability. This is particularly exemplified when considering so-called intra-rater and test-retest *reliability* studies. In general, *reliability* studies may capture one or more of the following sources of error/variability:

- Instrument reliability (*i.e. the reliability of the measurement device*).
- Rater reliability (*i.e. the reliability of the researcher/s or clinician/s administering the measurement device*).
- Response reliability (*i.e. the reliability/stability of the variable/construct being measured*).

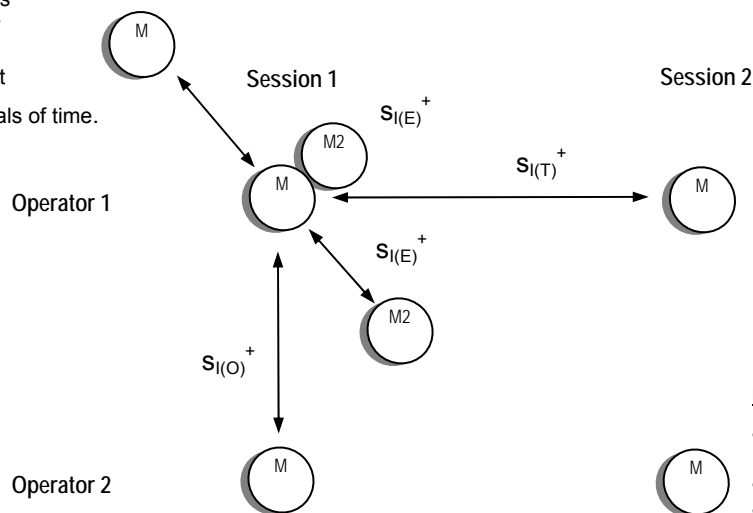
In most studies the isolation of rater from instrument and response *reliability* is often not possible. A study aimed at assessing the 'intra-rater' *reliability*, will not only capture the measurement error of the rater but also the stability of the construct measured and the measurement error of the assessment instrument. As such, it should be defined as a 'test-retest' study with the *reliability* of the measurement procedure as a whole under consideration rather than one of the potential individual sources of error or variability. The suffix term, to describe *reliability* studies is also a source of confusion and inconsistency (Table 10).

Although the concept of *reliability* would at face value seem simple, the nature of *reliability* theory can be complex with little consensus on the definition and most appropriate descriptive and conceptual terms. Perhaps the most standardised and robust definitions have been provided by the British Standards Institute (BSI). The BSI uses two terms, "trueness" and "precision" to describe what they term the accuracy of a

measurement method (ISO 5725-1, 1994). “Trueness” refers to the closeness of agreement between the test results and the true or accepted reference value. “Precision” refers to the closeness of agreement between test results where neither is considered a true value. The BSI identifies five different factors that may contribute to the variability of results (Figure 11). Three conditions of precision have been defined based on the factors that are allowed to vary between measurements (Figure 11). Where all factors are considered constant and the measurements are repeated within a short interval, the study is defined as repeatability (ISO 5725-1, 1994). Where all factors are varied with a larger time interval between measures, the study is defined as reproducibility (ISO 5725-1, 1994). Where one or more of the factors are allowed to vary between measurements, the precision is defined as “intermediate” (ISO 5725-3, 1994). The typical *reliability* study designs defined earlier are incorporated within the graphically representation of intermediate ‘precision’ shown in Figure 11 and Figure 12.

Repeatability:

- same measurement method
  - identical subjects
  - same laboratory
  - same operator
  - same equipment
- within short intervals of time.



Reproducibility:

- same measurement method
- identical subjects
- different laboratories
- different operators
- different equipment

**Figure 11.** Diagrammatic representation of BSI definitions for the accuracy of a measurement method for measurements on a continuous scale that yield a single value (devised by Webb and Rix, 2005, used with permission) (note: the single value may be the outcome of a calculation from a set of observations).

M – Measurement method; +Intermediate conditions of precision are specified as ‘precision conditions with M factor(s) different’, where M is the number of factors not maintained under constant conditions. The BSO ISO 5725 guidelines recommend the use of suffixes to specify which factors are not kept constant under intermediate precision conditions:

Operator                       $s_{I(O)}$  = operator-different intermediate precision standard deviation  
 Time                             $s_{I(T)}$  = time-different intermediate precision standard deviation  
 Equipment                     $s_{I(E)}$  = equipment-different intermediate precision standard deviation

If more than one factor is not maintained under constant conditions, it is specified as follows:

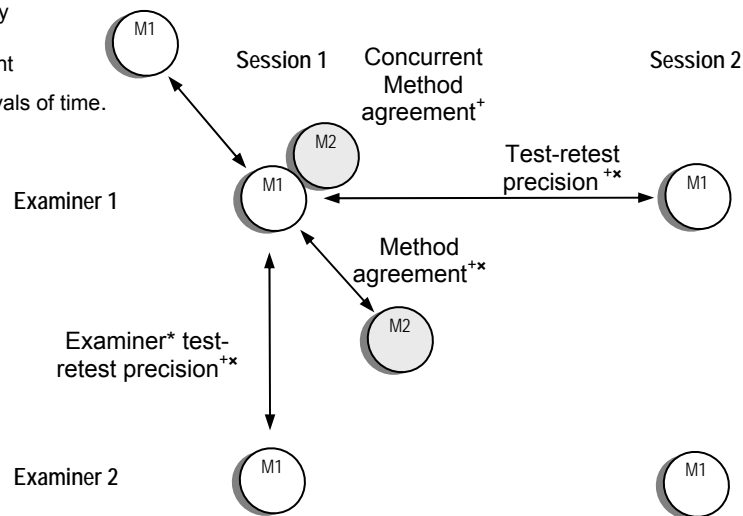
Example:  $s_{I(TO)}$  = [time + operator]-different intermediate precision standard deviation



Repeatability:

- same measurement method
- identical subjects
- same laboratory
- same operator
- same equipment

within short intervals of time.



Reproducibility:

- same measurement method
- identical subjects
- different laboratories
- different operators
- different equipment

**Figure 12.** Suggested terminology for common ‘Reliability’ studies using BSI framework (devised by Webb and Rix, 2005, used with permission).

M1 – measurement method 1; M2 – measurement method 2.

\*The term examiner is used interchangeably in the literature with rater, observer, tester, and operator.

+Intermediate levels of precision with the following factors selected as the variable:

Operator – examiner test-retest precision;

Time – test-retest precision;

Equipment, instrument or method- method agreement Note: encompasses same method but different equipment/instrument or a different method. An established method, equipment or instrument may be referred to as the ‘gold standard’ but this does not imply that it is measured without error.

×Biological variability is a limitation that, where relevant, is an additional variable that should be considered

Although ‘precision’ as defined by the BSI, and the intermediate conditions can be applied to many of the *reliability* studies seen in the biomedical literature, there are some limitations to the adoption of these terms and definitions: 1) Biological variation is not included in the factors listed as possible sources of variability. This variability may take the form of a learning effect, a change in the organism or an inability to repeat the task exactly a second time. Clearly, this biological instability is an important source of variation in biomedical research, particularly in the field of movement science and manual therapy where human performance measurements are commonly used. 2) Systematic bias is not included within the definition of precision. This is reflected in the

fact that with the BSI definition, precision is only measured in terms of standard deviation of the differences. Bias is only included in the definition of “trueness”. This form of quantification precludes a direct measure of systematic bias but may be affected by it (see later discussion). Systematic variability should be included in the initial assessment particularly if biological factors and examiner ratings are a considered a possible source of variation. 3) The terminology used with the intermediate levels of precision is not often utilised in biomedical literature and therefore likely to be quite unfamiliar to many researchers and clinicians. As such, for use in the biomedical arena, they have been modified as accurately as possible using more user friendly, familiar and recognisable terms. These terms are used throughout this thesis.

### **3.3 STATISTICAL APPROACHES FOR THE QUANTIFICATION OF AGREEMENT**

Irrespective of terminology used and the underlying study design, the most important aspect of any agreement study is ultimately the choice of statistical analysis and the inferences that are drawn from this. The goal of the statistical analysis is to detect and quantify disagreement, ultimately allowing an objective judgement on the acceptability of a test for daily clinical use. Improper use or faulty interpretation of the statistical parameters may result in invalid judgements on the acceptability of clinical tests (Westgard and Hunt, 1973).

A variety of different statistical approaches have appeared in the *reliability* literature. Although the definition and properties of each of these statistical methods is well established and there are an increasing number of appropriately worked examples and review papers in the literature, inappropriate analysis methods and interpretations continue to appear (White, 2004). Many of these approaches provide answers to the wrong questions, are used to provide answers to questions that they were not originally designed to answer, or have other limitations which significantly reduce the usefulness of the measure (Maher, 1993; Muller and Buttner, 1994).

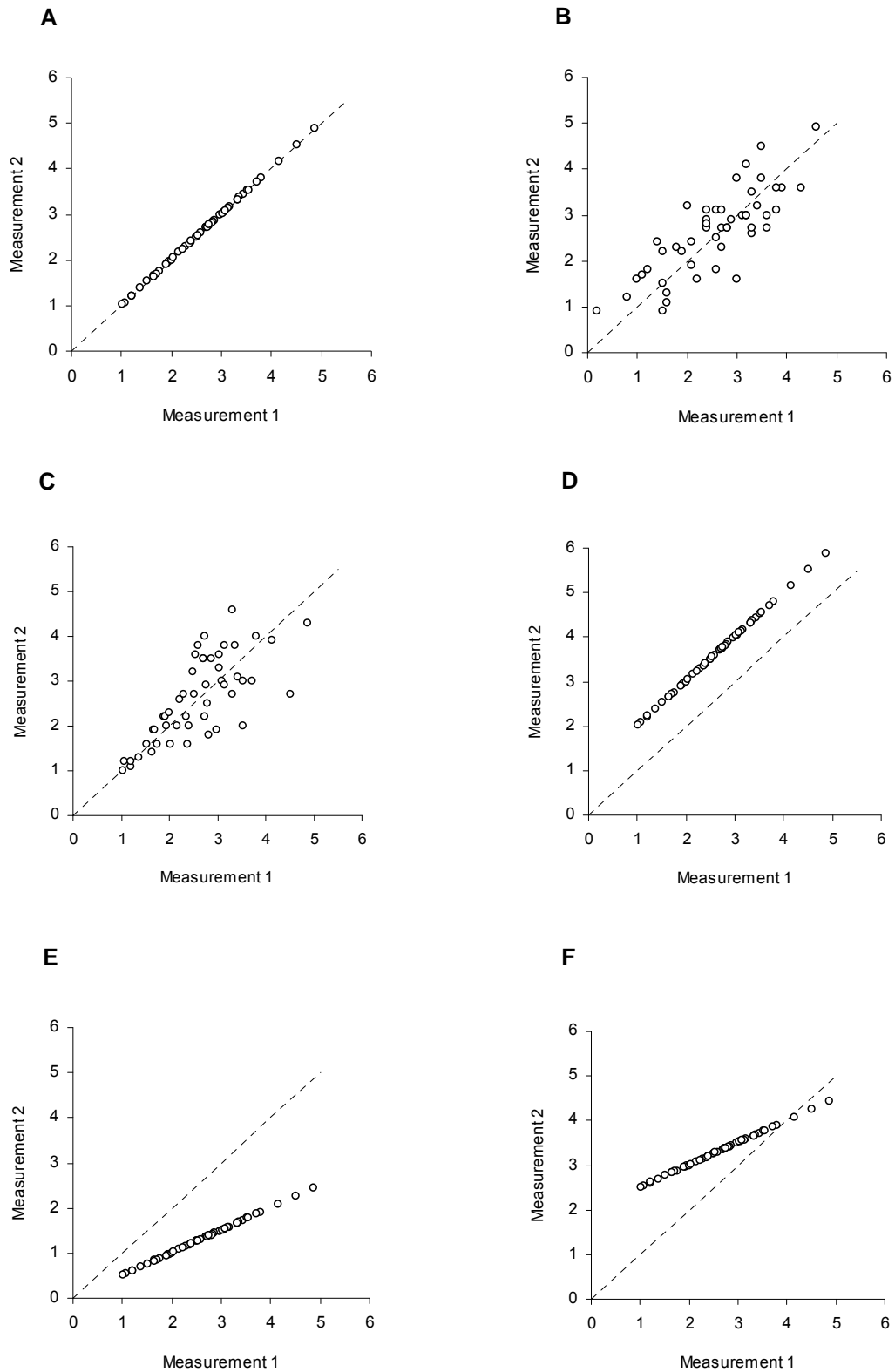
The inappropriate use of statistics may stem from the fact that despite the increasing number of helpful review papers, the analysis and interpretation of agreement studies is

one of the most confusing, controversial, and often impenetrable topics in the measurement and biostatistics literature, particularly to the non-statistician. The language can be tortuous, contradictory, inaccessible and sometimes very technical. Biostatisticians frequently disagree, sometimes sharply, on which statistical techniques are best for the appropriate analysis of method agreement and precision.

The following discussion focuses on the analysis of quantitative data which is measured on a continuous (interval) scale and normally distributed (parametric). To demonstrate the limitations and relative benefits of these approaches artificial, randomly generated data sets are used to represent HRA reference values. This simulation approach has been used by a number of authors to study the relative sensitivities of various *reliability* estimates (Westgard and Hunt, 1973; Bruton et al., 2000). Before discussing the relative merits of each statistical approach, it is worth introducing some key concepts and terms regarding the nature of agreement and variability that may be present between measurements. It is useful to know the types and magnitudes of errors and biases because each may have a different cause and affect the results in different ways.

When discussing the nature of agreement and the variability, it is often helpful to supplement this with a visual representation of the concepts (Figure 13). Using this approach, perfect statistical agreement can be represented using a scatterplot by all the data points falling exactly on a line marking a 45° angle (using same scale axes) and intersecting the axes at the origin (Figure 13A). This is often referred to as the “line of equality”. However, very few measurements can be made without some form of variability. Assuming the data sets are related in linear form, the total variability (total measurement error) can take two forms; error and systematic bias. The error component can be subgrouped into two types; *random* and *proportional*.

*Random error* (sometimes known as experimental error) represents the unpredictable component of repeated or concurrent measurements. *Random error* is typically assumed to be normally distributed with a zero mean and a constant variance. Using a scatterplot, *random error* is crudely reflected by the ‘tightness’ of a symmetrical clustering around the line of best fit (regression line) drawn through the data points. When no other bias is present, this data will cluster symmetrically around the line of equality (Figure 13B).



**Figure 13.** Graphical representation of systematic biases and error types between two measurements using scatterplots. A – no error or bias; B – random error; C – heteroscedasticity (proportional error) ; D – fixed (additive) bias; E – proportional (multiplicative) bias; F – proportional and fixed bias. The dotted line (----) represents the line of equality or perfect agreement.

*Proportional error* can be defined as a proportional (but symmetrical) increase in the *random error* across the distribution of measurements. It is distinguished by data points that progressively diverge or ‘fan out’ in a symmetrical fashion around the the line of best fit (Figure 13C). This is also known as *heteroscedasticity* and is commonly seen in measurement agreement studies (Ludbrook, 1997).

Systematic bias is defined as a bias which applies equally across the range of values in a data set. As with the definition of error, bias is often subgrouped into two types; *fixed* (constant or additive) or *proportional* (multiplicative). A *fixed bias* represents an additive (+ve or -ve) bias in one data set ( $Y = X + a$ ). Viewed graphically using a scatterplot, this is seen as an shift in the data set away from the line of best fit (change in intercept), with no change slope of this line of in the clustering of the data around the line (Figure 13D). A *proportional bias* represents a multiplicative difference ( $Y = bX$ ) in measurements between data sets (percentage or other proportional factor). This can be seen on a scatterplot as change in the slope of the relationship in differences with no change in the variance (Figure 13E). It is common for these errors and biases to occur in combination (Figure 13F).

An examination of the output parameters from the commonly used statistical analyses reveals that different parameters are sensitive to different type of error or bias (Table 11). Aspects of this have been demonstrated by a number of authors (Westgard and Hunt, 1973). Table 11 shows that some of these parameters are affected by only one error or bias type and provide a specific estimate of that error of bias (e.g., slope, Y intercept with regression). However, some parameters are affected by the variable but do not provide specific estimates (e.g., Pearson’s  $r$ ). In addition, many of the output parameters are affected by more than one error or bias type. When both sources of error are present in the measurements, the statistical parameter does not provide a specific estimate of either unless a modified analysis is performed.

These important issues will be developed further in the following sections when the benefits and limitations of each statistical approach are discussed in greater detail with examples using the artificial data set. Each section ends with a summary of the advantages and disadvantages of employing this particular approach in assessing agreement.

**Table 11.** Sensitivity of statistical parameters to different types of error and bias. Devised by Webb and Rix (2005), used with permission.

Type of Error/Bias	t-test (Paired)		Pearson's <i>r</i> Correlation	ICC		Simple Linear Regression			Bland & Altman	
	d	S <sub>d</sub>	<i>r</i>	Abs	Cons	Slope (b)	Y intercept (a)	S <sub>ε</sub>	d	1.96S <sub>d</sub>
Random Error		✓	✓	✓	✓			✓		✓
Fixed Bias	✓			✓			✓		✓	
Proportional Bias	✓	✓		✓	✓	✓		✓	✓ <sup>‡</sup>	✓ <sup>‡</sup>
Heteroscedasticity		✓	✓	✓	✓			✓*		✓ <sup>‡</sup>

*Abbreviations:* d – difference between measures; S<sub>d</sub> – standard deviation of difference; ICC – intraclass correlation coefficient; Abs – absolute; Cons – consistency; S<sub>ε</sub> – standard error of the estimate.

\* S<sub>ε</sub> will also be sensitive to heteroscedasticity unless a modified approach is used. This may involve a log transformation of the data or a weighted regression analysis

<sup>‡</sup> If a proportional bias or heteroscedasticity (proportional error) are present and the modified approach is not adopted (or data log transformed), d will be sensitive to a proportional bias and 1.96S<sub>d</sub> will be sensitive to proportional bias and heteroscedasticity – see text for further discussion. If a proportional bias present, a modified approach using simple linear regression provides a quantification of this bias type..

### 3.3.1 Pearson Product-moment Correlation (*r*).

The Pearson product-moment correlation (also referred to as the ‘Standard’ correlation coefficient, Pearson correlation coefficient or Pearson’s *r*) has often been used for analysis of precision, and method agreement studies. When this analysis is used, the investigator equates the degree of correlation (as indicated by *r*) directly with the level of agreement; the higher *r* (closer to 1.0), the greater the degree of agreement that is inferred. Intuitively, this statistical approach may seem quite attractive as it provides one simple index which on face value seems easy to interpret and apply. In addition, although Pearson’s *r* requires the data should be interval and normally distributed, relationships can be assessed between data sets with a different variance and units (i.e. cm's and degrees); as such it could be viewed as a bivariate *interclass* correlation coefficient (Schuck, 2004). However, there are numerous limitations to the use of Pearson’s *r* in the quantification of agreement. Indeed, this statistical approach has probably received the most critical attention from authors reviewing statistical approaches for measuring agreement.

Conceptually, Pearson’s *r* is a measure of the degree of association or covariation between two variables. It is a measure which gives information about the *relative* position within the two distributions rather than an *absolute* position relative to the

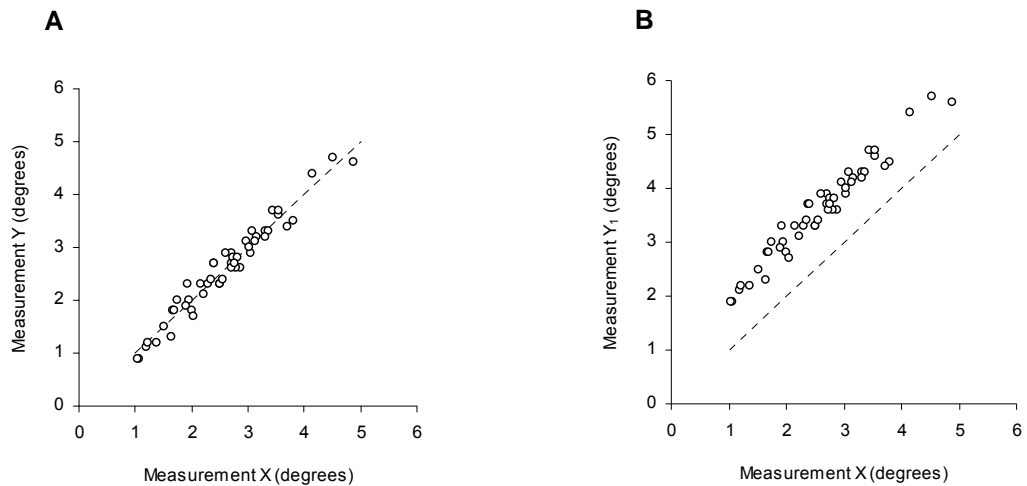
individual subject. Broadly speaking, as long as any variation is systematic (*fixed* or *proportional*) and therefore the ranking and relative position within the distribution stays the same between measures, Pearson's  $r$  will be relatively unaffected. Viewed from a more specific mathematical perspective, it captures the degree of *linearity* between two variables. In other words, it essentially describes how good a variable  $Y$  (second measurement) can be equated to another variable  $X$  (first measurement) by a transformation of the kind  $Y = bX + a$ ; an index of goodness-of-fit of a standard linear regression model to the observed values (Ludbrook, 1997). The result is that using Pearson's  $r$  alone, a linear relationship ( $Y = bX + a$ ) can be mistaken for good agreement. The injudicious use of the Pearson's  $r$  correlation has largely been attributed to a lack of clear understanding of this fundamental statistical property.

To illustrate the interpretation of correlation, limitations and how the results can be quite misleading, some numerical examples are given. The data used in these examples is data provided in Appendix II. The data was randomly generated using MINITAB<sup>®</sup> 12 statistical software (Minitab Inc., State College, Pennsylvania). The use of this data allows for known errors to be created and varied and as such, we know what the interpretation should be.

Fifty observations were randomly generated from a normal distribution with a defined mean and standard deviation (mean  $\pm$  SD:  $2.5 \pm 1$ ). This range and distribution was chosen so as to reflect the typical HRA-SSA measures seen in previous studies and is defined as degrees. These observations were duplicated to form the base observations for the  $X$  and  $Y$  data sets. A random measurement error was then introduced into the  $X$  and  $Y$  base observations. This was achieved by randomly generating two further data sets from a normal distribution with a defined mean and standard deviation (mean  $\pm$  SD:  $0 \pm 0.15$ ). One of these two sets of *random error* values were added to each of the base  $X$  and  $Y$  base observations. This gave artificial data representing two observations on each of a group of fifty subjects attended by a known measurement error. These observations could represent measurements from a method agreement, precision (test-retest of inter-rater) or repeatability study.

Figure 14A, shows the artificially generated data, with a line of equality ( $Y=X$ ). The correlation between  $X$  and  $Y$  is near perfect (Table 12; top row;  $r = 0.98$ ). Using this

result, the investigator would conclude that agreement is acceptable. A visual examination of the data shows that the measurements are closely related in a linear fashion (all data points cluster quite closely along a straight line), and there is little evidence of *fixed* or *proportional bias* (data lie very close to the line of equality).



**Figure 14.** Scatterplots of artificial measurements. A) XY data set – random error; B) XY<sub>1</sub> data set – random error and fixed bias. The dotted line (-----) represents the line of equality or perfect agreement.

**Pearson's *r* and fixed bias.** If measurement Y is transformed to data set Y<sub>1</sub> by adding 1.0 to all observations (Figure 14B), the equation for the relationship is now  $Y = X + a$  compared with  $Y = X$  for the previous example. However, when Pearson's *r* is calculated for XY<sub>1</sub>, the correlation is the same as Y with X (Table 12; rows 1 & 2;  $r = 0.98$ ); thus also suggesting the same level of agreement. The additive factor is reflected by all the data points lying above the line of equality; a clear systematic *fixed bias*. It is important to note that this insensitivity to any level of *fixed bias* means that very large, clinically meaningful biases of this type would be completely missed when using Pearson's *r*.



**Table 12.** Effects of various biases and errors on statistical results from Pearson's product moment correlation and Intra-class correlation coefficient analysis. Data is artificially generated with known errors and biases introduced

Data sets	n	Pearson's $r$		$r^2$	Intraclass Correlation Coefficients (ICC's)				
		$r$	P-Value		Absolute			Consistency	
					(1,1)	(2,1)	(3,1)	(2,1)	(3,1)
XY	50	0.98	< 0.0001	0.96	0.98	0.98	0.98	0.98	0.98
XY <sub>1</sub> (fixed) <sup>†</sup>	50	0.98	< 0.0001	0.96	0.50	0.60	0.60	0.98	0.98
XY <sub>2</sub> (proportional) <sup>‡</sup>	50	0.98	< 0.0001	0.96	0.41	0.53	0.53	0.89	0.89
XY <sub>3</sub> (fixed and proportional)	50	0.98	< 0.0001	0.96	-0.08	0.29	0.29	0.89	0.89
X <sub>1</sub> Y <sub>4</sub> (↑RE)	50	0.82	< 0.0001	0.68	0.83	0.82	0.82	0.82	0.82
XY <sub>5</sub> (heteroscedasticity)	50	0.74	< 0.0001	0.55	0.05	0.04	0.04	0.04	0.04
XY (range 0 to 2.5°)	25	0.94	< 0.0001	0.88	0.93	0.93	0.93	0.93	0.93
XY (range 2.6 to 5.0°)	25	0.95	< 0.0001	0.90	0.95	0.95	0.95	0.95	0.95
X <sub>1</sub> Y <sub>4</sub> (range 0 to 2.5°)	25	0.67	< 0.01	0.45	0.66	0.66	0.66	0.67	0.67
X <sub>1</sub> Y <sub>4</sub> (range 2.6 to 5.0°)	25	0.45	< 0.05	0.21	0.46	0.45	0.45	0.44	0.44
XY (half group)	25	0.98	< 0.0001	0.96	0.98	0.98	0.98	0.98	0.98
X <sub>1</sub> Y <sub>4</sub> (half group)	25	0.85	< 0.0001	0.73	0.86	0.86	0.86	0.86	0.86

Abbreviations: ↑RE – increased random error compared with XY data set; n – sample size;  $r^2$  – coefficient of determination

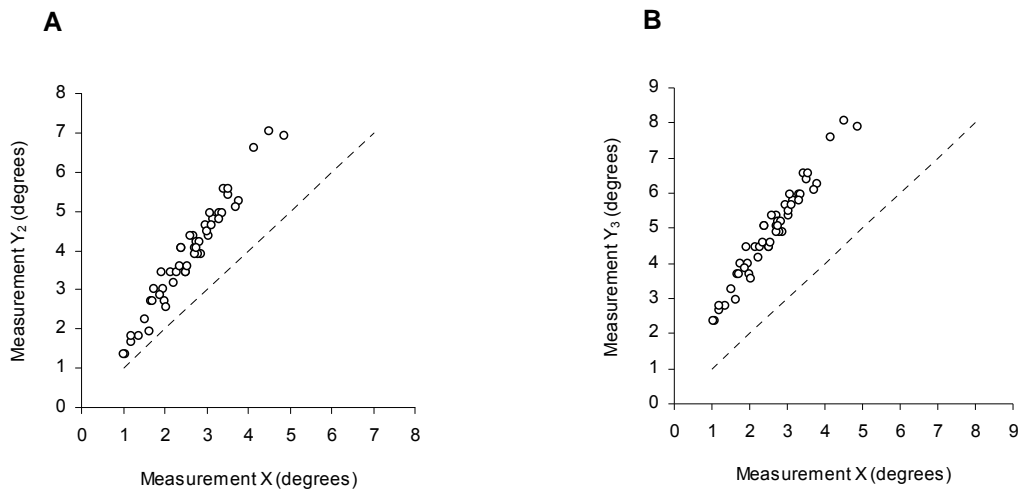
<sup>†</sup> Fixed (additive) bias of +1.0 degrees

<sup>‡</sup> Proportional (multiplicative) bias of 1.5 degrees

**Pearson's  $r$  and proportional bias.** Data set XY<sub>2</sub> represents the addition of a *proportional* (multiplicative) *systematic bias* to the XY data set (Figure 15A). In this case, the artificial measurements Y<sub>2</sub> were created by multiplying all the measurements in data set Y by 1.5; the equation for this relationship now being  $Y = 1.5X$ . The correlation using Pearson's  $r$  is unchanged compared with the XY data demonstrating the insensitivity of Pearson's  $r$  to a *proportional* (multiplicative) *systematic bias* (Table 12; row 3). If both the *fixed* and *proportional biases* are introduced to data XY representing the relationship  $Y = bX + a$  (Figure 15B; data set XY<sub>3</sub>; row 4), the correlation is again unaffected (Table 12; row 4).

From these examples it can be clearly seen that a high correlation coefficient ( $r$ ) does not necessarily imply close agreement. Good absolute agreement is only obtained when

the pairs of readings closely follow the line of equality. Pearson's  $r$  does not test against this line of equality. As long as the data points lie on a straight line, the Pearson's  $r$  indicates a perfect relationship between them (Schuck, 2004). As such it is effectively 'blind' to the presence of a *fixed* (additive) &/or *proportional* (multiplicative) *bias* (Bland and Altman, 2003). These 'unseen' biases may be very large, potentially having important scientific and/or clinical implications. For this reason, if a correlation analysis is going to be used, it is essential to first plot a scatterplot to look at the general trends and relationship of that data so that correct inferences can be drawn.



**Figure 15.** Scatterplots of artificial measurements. A) XY<sub>2</sub> data set – random error and proportional bias; B) XY<sub>3</sub> data set – random error and fixed bias. The dotted line (-----) represents the line of equality or perfect agreement.

**Pearson's  $r$  and random error.** The Pearson's  $r$  statistic is sensitive to the *random error* (or variance) between paired measurements. As stated earlier, using a scatterplot this is crudely reflected by the 'tightness' of clustering around the line of best fit drawn through the data points. To demonstrate the effect of any increased *random error*, a set of observations (Table 12; data set XY<sub>4</sub>; row 5) were generated in the similar fashion to

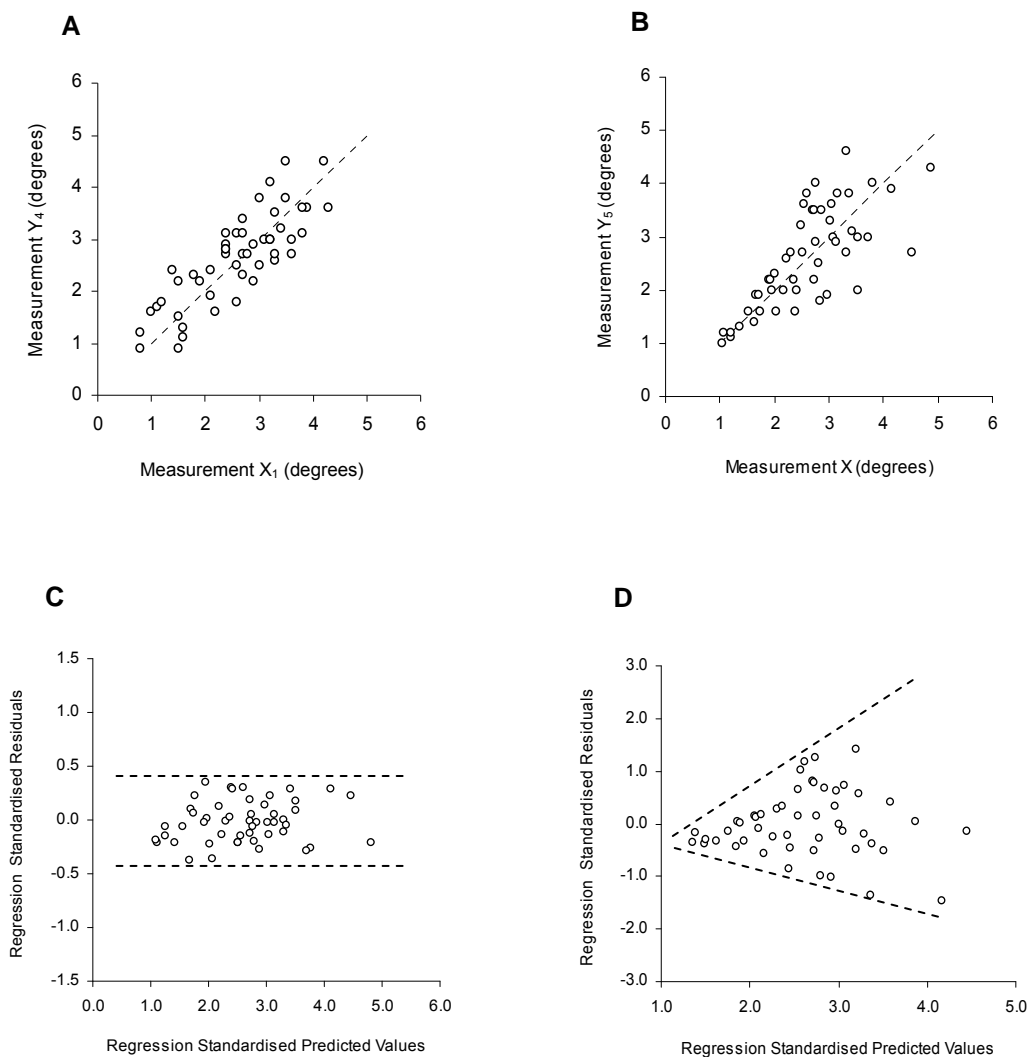
the original XY measurements but with triple the standard deviation for the error ( $0.45^\circ$  vs  $0.15^\circ$ ). Figure 16A shows a scatterplot of the data ( $XY_4$ ). When compared with the original XY data (Figure 14A), there is a clear visual increase in size of the cloud of data points around the line of best fit, but no alteration in the linear relationship ( $Y = bX + a$ ). The correlation for  $X_1Y_4$  is now less than for XY (Table 12;  $r = 0.82$  vs  $0.98$ ). When  $r$  is used as an index of agreement, this suggests a lower level of agreement for data set  $X_1Y_4$  compared with XY; a reasonable inference in this instance.

***Pearson's r and Heteroscedasticity.*** The data sets in all the correlation examples so far exhibit a linear relationship and *homoscedasticity*; that is, the relationship between them appears linear in nature and the variance in differences between measurements is the same (homogenous) across all the measurements. As highlighted earlier, if the variance in differences changes across the distribution of measurements, then there is *heteroscedasticity* in the data. In most cases, the variance increases with measurement size. The typical result is a 'funnel' shape array of data points as shown in Figure 16B (data set  $XY_5$ ). Pearson's  $r$  is sensitive to this variance characteristic (Table 12; row 6); the more heteroscedastic the data, the lower the  $r$  becomes. This effect is demonstrated with a lower correlation between X and  $Y_5$  compared with that for X and Y (Table 12;  $r = 0.74$  vs  $0.98$ ). The presence of *heteroscedasticity* can be checked more rigorously with a plot of the regression standardised residuals against the regression standardised predicted values. The residuals are the difference between the values of the outcome predicted by the model and the values of outcome observed in the sample; in this sense they represent the error in the model (Field, 2005). Graphically, the residual value is the vertical, horizontal or perpendicular distance of the data point from the regression line (Figure 21A; page 98). The standardised values are the residual values divided by an estimate of their standard deviation (Field, 2005). If the points on the scattergraph appear randomly and evenly dispersed with no 'funnelling' then homoscedasticity exists (Figure 16C). If the graph funnels out, then the chances are that there is *heteroscedasticity* in the data (Figure 16D).

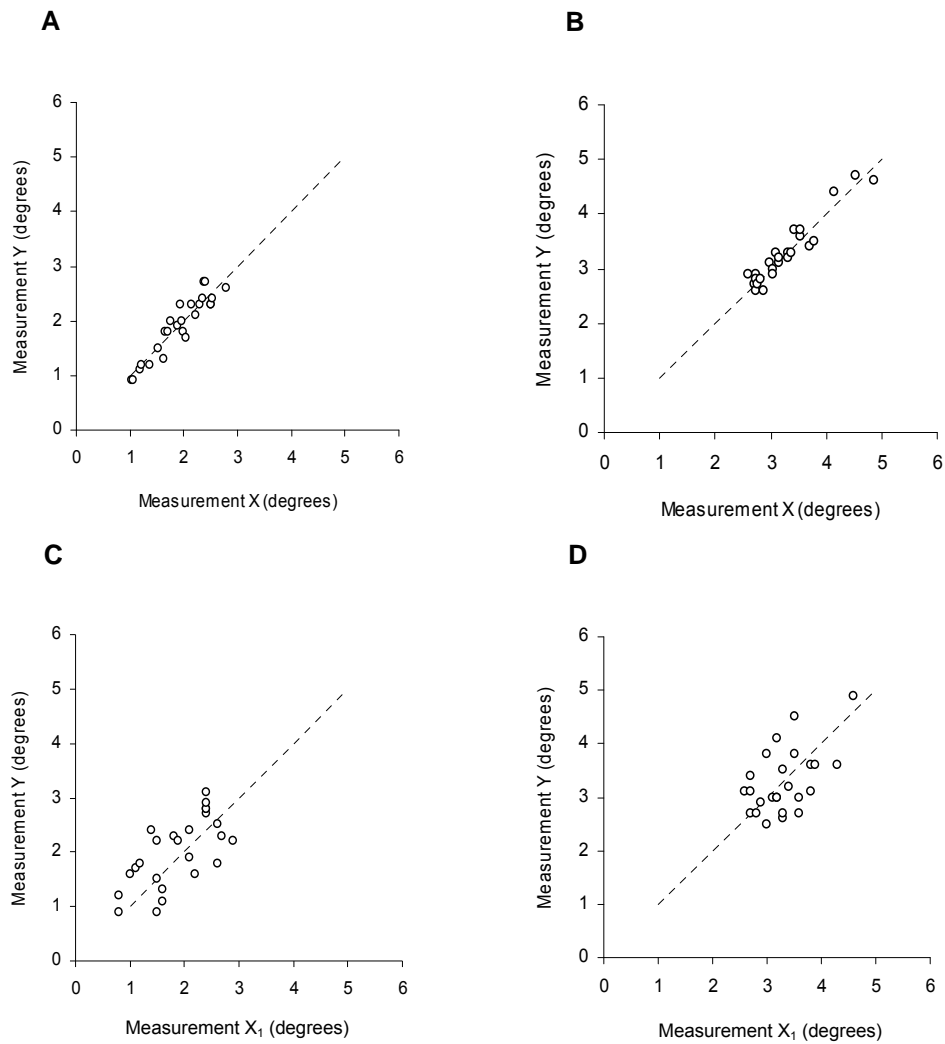
***Pearson's r and range of values.*** A further limitation of Pearson's  $r$ , and correlation coefficients in general, is the dependence on range of the measurement values (heterogeneity) in the study. This limitation becomes more apparent as *random error*

increases relative to the range. To demonstrate this, the data set XY and  $X_1Y_4$  (increased *random error*) were split in half to produce high and low range data sets (Table 12; rows 7 to 10). The plots of these data sets are shown in Figure 17A-D.

The correlation coefficients ( $r$ ) for both high and low range values for the XY data set are relatively unaffected (Table 12; rows 7 & 8). However, for both the restricted range data sets with the larger *random error* ( $X_1Y_4$ ), they are noticeably less than  $r = 0.82$  for the whole  $X_1Y_4$  data set (Table 12; rows 9 & 10).



**Figure 16.** Scatterplots of artificial measurements. A)  $X_1Y_4$  data set – increased random error; B)  $XY_5$  data set – heteroscedasticity; C) & D) residuals against predicted values (regression standardised) for data sets  $X_1Y_4$  and  $XY_5$  respectively. The dotted line for A) & B) (-----) represents the line of equality or perfect agreement.



**Figure 17.** Scatterplots of artificial measurements. A) XY data set – low range ; B) XY data set – high range; C) X<sub>1</sub>Y<sub>4</sub> data set – low range ; D) X<sub>1</sub>Y<sub>4</sub> data set – high range. The dotted line (-----) represents the line of equality or perfect agreement.

It might be argued that halving the sample size ( $n = 25$  from  $n = 50$ ) contributes to the effects shown with these examples. However, if number of measurements for data sets XY and X<sub>1</sub>Y<sub>4</sub> are halved by removing alternate data points, the correlation coefficients remain relatively unaffected (Table 12; rows 11 & 12; XY,  $r = 0.98$ ; X<sub>1</sub>Y<sub>4</sub>,  $r = 0.85$ ).

These examples demonstrate that provided the *random error* is sufficiently large, the sampling theory (sample chosen) in the study can have a meaningful effect on the Pearson's correlation coefficient. The higher the range of values in that population, the higher the coefficient; if the population has a different range than the representative population, then the coefficient will be different (Bland and Altman, 2003). Therefore, the correlation coefficient potentially only has meaning for the population from which the study subjects can be regarded as a random sample (Altman and Bland, 1983; Bland and Altman, 1996a; Bland and Altman, 2003). If Pearson's  $r$  is applied to data from a group of individuals demonstrating a wide range of the measured characteristic, precision may appear higher than when applied to a group demonstrating a narrow range of the characteristic. Direct comparisons are therefore only valid when comparing like populations (range of measurements in particular) and should only be used if we have a representative sample of the subject population we wish to study (Bland and Altman, 2003). From another perspective, it has been suggested that correlation coefficients should only be used with a fixed population that can be well defined (Quan and Shih, 1996). When looking at method agreement it is conceptually wrong to have a level of agreement that changes with the range of measurements.

***Pearson's  $r$ : Quantifying error/bias size & level of agreement.*** An important limitation of Pearson's  $r$  (and any *reliability* coefficient) in quantifying agreement, is the indirect nature of the index. As this index does not quantify the size of error in the actual scale of measurement used, a judgement cannot be made on whether the error is clinically acceptable for a given purpose. Take the measurement of diastolic blood pressure (BP) as an example. We could decide that a 10mmHg change will affect clinical decision making regarding patient management. An instrument reading that varies by as much as 20mmHg on repeated measurement (assuming stability of the BP), would therefore not be suitable for monitoring changes in diastolic BP. We would not be able to judge whether a 10mmHg change in diastolic BP was real, or due to the relative measurement precision. The correlation coefficient does not tell us whether the error is 5, 10, or 20mmHg. There could be a high correlation coefficient but an actual range of error/variability which means that the test would not be suitable for this purpose. In the artificial data shown in Figure 14A, the relationship closely matches the line of equality and  $r = 0.98$ . But what is the actual size of *random error*,  $0.1^\circ$ ,  $0.5^\circ$  or  $1.0^\circ$ ? This is not given by the Pearson's  $r$  output parameters and only a crude visual estimate can be

obtained from scatterplot. As such,  $r$  does not give us an index of agreement that allows the assessment of suitability for a given diagnostic task and also that can be directly applied to any given subject/patient in the clinical environment.

***Pearson's  $r$  and statistical significance.*** When using the Pearson product-moment correlation for assessing agreement, some authors use the statistical significance of the correlation ( $P$ -value) as a secondary indicator of the level of agreement. The  $P$ -value within a Pearson's statistic, tells us the probability of the correlation occurring by chance. From the outset, there is a huge conceptual leap between this hypothesis test and what is required for a useful agreement measure. If it were to be used as an indicator of agreement, then clearly a one-tailed hypothesis test must be selected as a specific direction of the hypothesis is being tested (positive correlation). However, the test of statistical significance is not a good guide to the 'real' significance of the correlation. With large sample sizes, the value of  $r$  required to achieve statistical significance (i.e., to show that there is some relationship between the two variables) is rather low. It is perhaps better to use the value of  $r^2$  (coefficient of determination) as an indicator of the real significance as this value shows the amount of variation in one variable explained by the other. This is demonstrated in Table 12 using the artificial data. For most data sets, the  $P$ -values are consistent and highly significant indicating that the probability of the correlations occurring by chance is extremely low in most cases. However,  $r$  does vary across the data sets and the amount of variation that can be explained in the Y measurements by a change in X ( $r^2$ ) varies from 90% to 21%. This will be discussed further in the following section on regression analysis.

### **Summary: Pearson's $r$ as a measure of agreement**

Overall then, the use of Pearson's  $r$  as a measure of agreement should be discouraged. Conceptually, this statistical approach examines the wrong premise and is only sensitive to *random error* (or variance). Without a prior exploration of the measurements using a scatterplot and comparison with the line of equality, inferences on the level of agreement can be highly misleading. It has been argued that Pearson's  $r$  may have some value when there is no systematic bias (*fixed &/or proportional*) or when only the *random error* (or variation) is of interest (provided there is no *heteroscedasticity*). However, the indirect nature of the index (size of agreement not quantified in scale of

measurement used) and the dependence on the range of values, significantly limit the usefulness of this application particularly where method agreement studies are concerned.

### **3.3.2 Intraclass correlation coefficients ( $r_i$ )**

The intraclass correlation coefficient (ICC) is an alternative to Pearson's  $r$  when one is interested in assessing the relationship between data sets with a common variance and metric (common *class*). Like Pearson's  $r$ , the ICC is a single index but is calculated using variance estimates obtained through the partitioning of total variance into between and within subject variance (analysis of variance or ANOVA). As with Pearson's  $r$ , the degree of correlation is indicated by on a scale of 0 to 1.0 (although negative values can be produced); good agreement being equated with high correlation coefficients. This statistic was developed at the turn of twentieth century to measure concordance of genetics essentially using a one-way ANOVA model. Thereafter, the ICC gained entry first into psychology, mainly for test-retest studies and then into general medicine. These ICC's are more complicated and were developed using a two-way ANOVA model. Over the years, several versions of the ICC have been described. The most commonly used are derived from Shrout and Fleiss (1979) and Fleiss (1986); these versions will form the basis of this discussion. The application of ICC's have generally been well described for inter-observer precision studies but less so for use with test-retest precision, repeatability and method agreement studies. Several authors have reviewed the application of this statistic with varying degrees of detail, insight and clarity; sometimes with useful worked examples (Lee, Koh and Ong, 1989; Bland and Altman, 1990; Haas, 1991b; Muller and Buttner, 1994; McGraw and Wong, 1996; Atkinson and Nevill, 1998; Rankin and Stokes, 1998; Bruton et al., 2000; Hopkins, 2000; Schuck, 2004; White and Van Den Broek, 2004). However, some confusion and disagreement still exists, certainly in the movement and manual therapy sciences, concerning both which one to use in a given situation, and how to calculate and interpret the chosen ICC.

***ICC models.*** Before discussing the limitations of ICC's for the analysis of agreement studies it is first worth briefly outlining the ICC models and decision faced when choosing an ICC. As already stated, these models are derived from one-way or two-way



repeated measures ANOVA. A convenient way to view the potential sources of variance is shown in Table 13. In this table,  $i$  is used as the subscript for the randomly chosen objects (often called targets) of measurement (which vary in number from 1 to  $n$ ). The observations on each object of measurement are represented in the columns of the table using  $j$  as the subscript (which vary in number from 1 to  $k$ ). These observations may be made using different methods (method agreement), raters (inter-rater precision) or repeated measures using the same rater and instrument (repeatability or test-retest precision). Both  $i$  and  $j$  are also referred to as factors. Because the objects are randomly selected, they represent a random factor in the design. This ‘row’ factor is common to all ICC models. What differs between models is the ‘column’ factor ( $j$ ).

Object of measurement	Measurement			
	1	2	... $j$	... $k$
1	$X_{11}$	$X_{12}$	... $X_{1j}$	... $X_{1k}$
2	$X_{21}$	$X_{22}$	... $X_{2j}$	... $X_{2k}$
.	.	.	.	.
.	.	.	.	.
$i$	$X_{i1}$	$X_{i2}$	... $X_{ij}$	... $X_{ik}$
.	.	.	.	.
.	.	.	.	.
$n$	$X_{n1}$	$X_{n2}$	... $X_{nj}$	... $X_{nk}$

**Table 13.** A convenient data matrix, notational system and source of variance (factors) for data used in calculating Intraclass correlation coefficients.

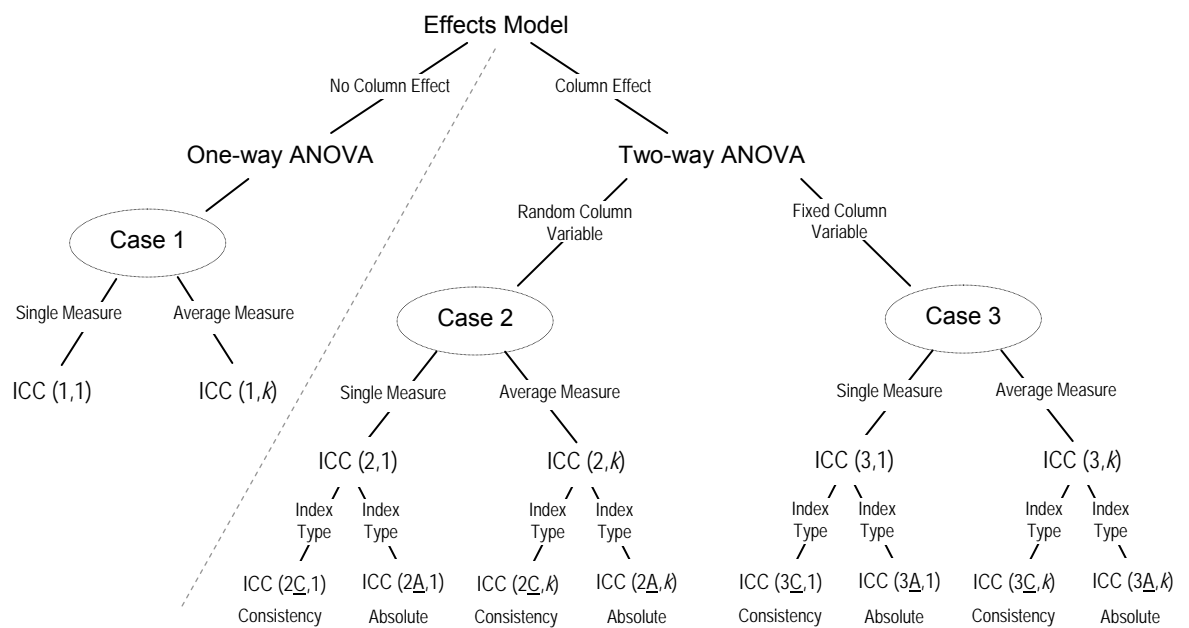
If within the study design, it is believed that  $j$  (different measures – column data) is not a source of variance (error) and that the ordering of data collection is irrelevant, then a one-way ANOVA model would be appropriate; i.e., only one factor (row data) is considered as a systematic cause of differences in the data. A commonly applied test for this ‘interchangeability’ of the column data suggests that for the column data to be truly irrelevant as a systematic source of variance, half the data from one column could be interchanged with another with no effect on the statistical outcome. For studies in which a source of systematic variance is believed to be associated with the column data ( $j$ ; measures) as well as the row data ( $i$ ), a two way ANOVA model would be used to define the ICC.

The two-way ICC model can be further subdivided based on whether the column factor is truly *random* or represents a *fixed* effect. For instance, if two raters (or measurement instruments) have been randomly drawn from the total population of raters (or available instruments) then a change in the raters should have no effect on the statistical probabilities in the study. If the column variable is not a random selection of all possibilities and changing the variable will likely have an effect, the column factor is considered *fixed*. Since the object of measurement is always a random factor, we have a *mixed* model if we assume the *fixed* measurement error (column variable) and a *random* model if both factors are random. The distinction between *random-fixed* effects does not affect the calculation of the ICC, only the interpretation (Table 12). Namely, when levels of the column factor are randomly sampled, one can generalize beyond one's data, but not when they are fixed. The three ICC models that result from this reasoning on sources of variation and sampling theory, are commonly referred to as a *Case 1 (A)* for the one-way model, and *Cases 2 (B)* and *3 (C)* for the two-way *random* and *mixed* models respectively (Figure 18).

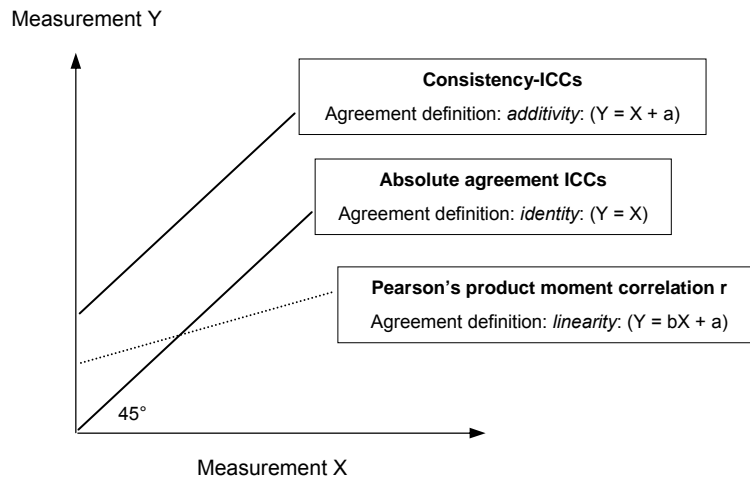
Within each *Case*, several more ICC variations are possible (Figure 18). The first of these are the *Type 1* and *Type k* ICC's. These relate to the number of measures used to derive the value used in the ICC analysis. In situations where the value for each measurement object (subject) is derived from a single measure, a *single measure (Type 1)* ICC should be applied. When the value for each measurement object represents the average of *k* measures (e.g.; average of *k* repeated readings, the average of *k* raters scores or the average of a *k* item test), an average measure (*Type k*) ICC should be applied. Strictly speaking, ICC's for single scores are just a special case of the ICC's for average measures, and they tend to yield lower values than the corresponding average measures ICC.

Within the two-way ICC models (*Cases 2* and *3*), the two further ICC choices are available; the *absolute* and *consistency* ICC's. These are based on type of agreement definition used and are of great importance, both computationally and inferentially. Perfect absolute agreement is represented on a scatterplot by all the data points falling exactly on a line making a 45 degree angle and intersecting the axes at the origin in other words, the data would sit exactly on a line of equality (Figure 13A). In this context, an *absolute* ICC captures this degree of *identity* ( $Y = X$ ) between two

measurements (Figure 19). The *consistency* ICC is an index of *additivity* reflecting how good the second measurement variable (Y) can be equated from the first (X) by a formula like  $Y = X + a$ . This arises from the fact that the between-measures variance is excluded from the denominator in the consistency ICC formula. It should be pointed out that the ICC's for Case 1 (one-way ANOVA) are both based on *absolute* agreement definitions. The importance of these *absolute* and *consistency* ICC variations will be discussed shortly when the limitations of ICC's as measures of agreement are considered.



**Figure 18.** Flow chart for selecting an appropriate intraclass correlation coefficient (ICC). Adapted from McGraw and Wong (1996)



**Figure 19.** Graphical representation of agreement definitions in various correlation coefficients. ICC = Intraclass correlation coefficient (adapted from Schuck, 2004)

**Properties of the ICC.** The ICC is commonly used for the analysis of precision, repeatability and method agreement studies. As a simple index, it shares the same intuitive attractiveness as Pearson's product moment correlation and as with Pearson's  $r$ , the investigator equates the degree of correlation directly with the level of agreement; the higher the ICC value the closer the degree of agreement that is inferred. When the data sets share a common *class*, the ICC has distinct advantages over Pearson's  $r$ , the most obvious being that more than two data sets can be compared. In addition, within the previous section, it was shown that Pearson's  $r$  is sensitive to *random error*, but 'blind' to *proportional* and *fixed bias*. As a consequence, a linear relationship ( $Y = bX + a$ ) can be mistaken for good agreement and therefore inferences on the level of agreement can be highly misleading. This can be partly or completely avoided using ICC's as a measure of agreement.

All ICC's are sensitive to *random error* (Table 11). An ICC with a *consistency* definition of agreement ( $Y = X + a$ ; *additivity*), is also sensitive to *proportional bias* but is still 'blind' to (or ignores) the *fixed bias*. An ICC with *absolute* definition of

agreement ( $Y = X$ ; *identity*), is sensitive to both *proportional bias* and *fixed bias* (Table 11 & Figure 19).

These properties are nicely highlighted with the artificial randomly generated data described in the previous subsection (Table 12). When data set XY is considered (Table 12; row 1), the correlation between X and Y is near perfect with all the ICC models. However, when a *fixed bias* (additive or constant) is introduced (Table 12; row 2; data set XY<sub>1</sub>;  $Y = X + a$ ), the *absolute* ICC values are notably reduced whilst the *consistency* ICC values are unchanged reflecting the respective agreement definitions. A *proportional* (multiplicative) *bias* (Table 12; row 3 data set XY<sub>2</sub>;  $Y = bX$ ) results in a reduction in the ICC values compared with the XY data for both agreement definitions but, interestingly this is greatest for the *absolute* variation. A combination of biases (Table 12; row 4; data set XY<sub>3</sub>;  $Y = bX + a$ ) has a marked effect on the *absolute* variations but leaves the *consistency* variations unchanged compared with data set XY<sub>2</sub> (*proportional bias*). The artificial data also demonstrates that as with Pearson's *r*, all ICC models and variations are sensitive to the *random errors* (or variance) between paired measurements and *heteroscedasticity* in the data (Table 12; rows 5 & 6; data sets X<sub>1</sub>Y<sub>4</sub> and X<sub>1</sub>Y<sub>5</sub>). However *heteroscedasticity* has noticeably more marked effect on the ICC values compared with Pearson's *r* further highlighting the different inferences that may be generated from these statistical approaches (Table 12).

***The ICC and identification of error/bias types.*** Although the ICC has distinct benefits over Pearson's *r* when assessing agreement between measures there are still some significant limitations to the use of intra-class correlation coefficients in the quantification of agreement. Firstly, from the previous section, it should be clear that inferences may be misleading if the distinction between a *consistency* and *absolute* ICC model is overlooked; most significantly a *fixed* (constant; *additive*) bias will be missed when using a *consistency* agreement definition ICC. In addition, although the ICC provides an increased sensitivity to systematic bias compared with Pearson's *r*, when a difference between data sets exists, the ICC statistical analysis does not reveal the type of error or bias present and the relative impact of these on the relationship observed. Put another way, a low *absolute* ICC value considered in isolation from any other analysis, would not tell you whether the poor correlation was due to *random error* and/or *systematic bias* (*proportional* and/or *fixed*). Identification and quantification of the bias

and error types may affect the inferences made from the study and the direction of further investigation. It has been argued that if for a given comparison of data, there is no difference between *consistency* and *absolute* ICC variations, any differences cannot be attributed to a *fixed bias*. However, the results shown in Table 12 demonstrate that this may not always be the case when a *proportional* (multiplicative) *bias* is present (row 4; data set XY<sub>3</sub>). As the following subsections will reveal, there are more direct and accurate ways of assessing for the presence of *fixed bias*.

**The choice of ICC.** Another potentially significant problem with the use of ICC's surrounds the choice of ICC. As stated earlier the most commonly used ICC equations are derived from Shrout and Fleiss (1979). McGraw & Wong (1996) added the *consistency & absolute* variations to Cases 2 & 3 resulting in ten ICC variations. These models and calculations cover most study scenarios faced in the manual therapy and biomechanics literature and the computational formulae have also found themselves into SPSS statistical software from version 8.0. The choice of ICC can be quite perplexing (Weir, 2005). This can be a particular problem with test-retest & method agreement studies as much of the literature deals with inter-observer precision. However, the choice of ICC for each scenario should be relatively straight forward if the conceptual framework of each model & equation is clear; the reasoning behind each step in the selection process is understood; and this is tied in with good 'reliability' theory, a clear understanding of the nature of error under question and the potential application of the results. The potential misleading impact of *consistency vs absolute* ICC variations has already been discussed. However, problems can arise at an earlier stage in the selection process.

From the analysis of artificial data (Table 12), it can be seen that the presence of a *proportional* and/or *fixed bias* (rows 2-4; data sets XY<sub>1</sub> – XY<sub>3</sub>) has a greater effect on the Case 1 ICC (ICC (1,1)) than the corresponding absolute ICC's derived from Cases 2 and 3. The relatively 'conservative' nature of Case 1 ICC's has been highlighted by several authors (Haas, 1991b; Muller and Buttner, 1994; Schuck, 2004). So as with the *consistency vs absolute* ICC variations, the choice of ICC model between Case 1 and Case 2/3 could affect the inferences on the levels of agreement between measures. The choice between ICC's derived from one-way ANOVA (Case 1) vs two-way ANOVA

(Cases 2 & 3) seems to cause the most difficulty with test-retest precision study designs (Bland and Altman, 1996b; Schuck, 2004).

When discussing this issue, Schuck (2004) highlights that the difference between the *absolute* ICC's is dependent upon the mean difference between the measures. At low mean differences, the one-way ICC's tend to be higher than their two-way *absolute* counterparts. As the mean difference increases, both ICC's reduce in magnitude but this is greater for the one-way ICC's. As a consequence, there is a 'cross-over' point beyond which the two-way ICC's are relatively larger. It should be pointed out that for most data sets, the computational differences between the two ICC models is usually only small.

The potential to select the wrong equation is increased further if the ICC versions outside those derived from Shrout and Fleiss (Shrout and Fleiss, 1979; Fleiss, 1986) are considered. The various ICC approaches can yield hugely different results thus leading to radically different conclusions about the level of agreement between measures (Haas, 1991a; Muller and Buttner, 1994; Schuck, 2004).

***Further limitations of ICCs: Range of data set values, statistical significance & quantifying error/bias size & level of agreement.*** Some of the remaining limitations are the same as those described for Pearson's  $r$ ; namely that the ICC value obtained is dependent on the range (heterogeneity) of the variables (Table 12; rows 7-10), does not quantify the size of error in the actual scale of measurement and no use for  $P$ -value in assessing agreement. The reader is referred back to the Pearson's  $r$  section for review of these limitations. With respect to the influence of heterogeneity, this can be explained if the ICC is viewed in simple terms as the ratio of true score variance (between-subjects variance) to true score variance plus error. If true score variance is sufficiently large, agreement will always appear high and vice versa. Following from this, it should be possible to see that a large ICC can mask poor agreement when the between-subject variability is high. Conversely, a low ICC can be found even when between measures variability is low (good agreement) if the between-subject variability is similarly low.

In addition to the issues surrounding the scale of measurement, the limitations discussed in the Pearson's  $r$  subsection also highlighted the difficulty in judging what magnitude

of correlation coefficient represents acceptable agreement. Some sources have attempted to delineate good, medium and poor levels for the ICC, but there is no consensus as to what value of ICC represents acceptable agreement. It has been suggested that a cut-off point for minimum acceptability could be determined by a comparison of the ICC to a minimum that would be acceptable ( $ICC > 0.6$ , Eliasziw, Young, Woodbury and Fryday-Field, 1994). In the context of method agreement studies, it has also been recommended that the lower limit of the CI for the ICC should be at least 0.75 for agreement to be considered acceptable (Lee et al., 1989). However, this is only one of three statistical conditions that they suggest are met before two methods can be judged as interchangeable. It is clear that most suggestions for cut-off points are at best arbitrary with little scientific or statistical underpinning. It is not theoretically defensible to set a universal standard for agreement (e.g.,  $ICC > 0.75$ ) when the value depends on the version of ICC used and the range (heterogeneity/between-subject variance) of the data.

Another problem is that the ICC values may be negative whereas the corresponding parameters and the slope of the relationship are strictly positive. How such negative values should be interpreted is quite unclear (Muller and Buttner, 1994). Although not unique to the ICC, a further limitation relates to the underlying statistical assumptions for ICC analysis. As mentioned earlier, the ICC's are calculated from repeated measures ANOVA using data sets with a common variance and metric (common *class*). However, the output from repeated measures ANOVA is particularly affected when the assumption of homogeneity of variance is not met. In this instance, Lin's concordance correlation coefficient (absolute agreement) is recommended (Schuck, 2004).

### **Summary: The ICC as a measure of agreement**

In summary, the ICC has distinct benefits over Pearson's  $r$  when assessing agreement between measures. However, there are still some significant limitations to the use of intra-class correlation coefficients in the quantification of agreement that make the use of this statistic highly questionable. If the ICC is utilised for measuring agreement, the investigator should clearly report which ICC was used (including the equation) and reasons for choice (Atkinson and Nevill, 1998; Schuck, 2004). To further aid comparison between studies, the range of measurement variables (between-subject variation) in the study should be reported; this could be done by reproducing the



ANOVA table variance parameters or simply the range of the measured characteristic. The ICC should never be used in isolation without a prior exploration of the measurements using a scatterplot and comparison with the line of equality. The ICC should ideally be used in conjunction with agreement measures that allow a quantification of the nature and size of random error and systematic bias present in the study. As with Pearson's  $r$ , the use of ICC's in method agreement studies should be discouraged.

### 3.3.3 Simple Linear Regression

Simple (ordinary or standard) linear regression is a very widely used statistic in biology and medicine. However, in the arena of agreement studies, it is probably one of the most inappropriately utilised statistical approaches after Pearson's  $r$ . Investigators often use this statistical approach in addition to a correlation coefficient as they feel it provides more robust and complete assessment of agreement. Indeed, this approach is sensitive to all the main error and bias types seen with agreement studies unlike Pearson's  $r$  (Table 11). In addition, the output parameters do provide an individual quantification of the magnitude of these errors and biases in the scale of measurement used. Therefore, intuitively, if there is a linear relationship between the two measurements (as in most agreement studies), it may seem logical to use linear regression analysis when assessing agreement. However, as with Pearson's  $r$ , the use of these regression analytical techniques can lead to flawed conclusions. From the outset though, it should be said that the use of regression statistics is not totally unsuitable for the quantification of agreement although some applications of simple linear regression are inappropriate to all three study designs (test-retest precision, inter-examiner precision & method agreement) under consideration (Bland and Altman, 2003). Most of the problems arise from the inappropriate use of the ordinary least squares (OLS) linear regression analysis (Ludbrook, 1997; Ludbrook, 2002; Bland and Altman, 2003).

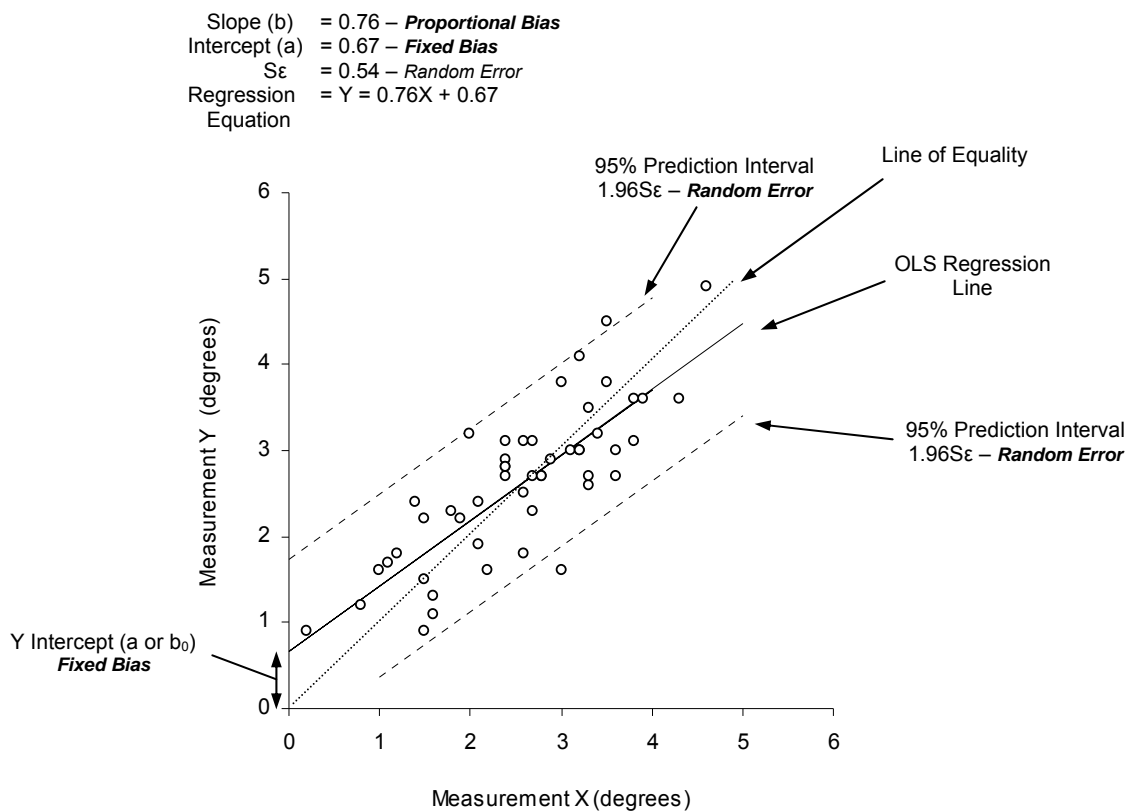
***Properties of simple linear regression and identification of error/bias types.*** In discussing the limitations of simple linear regression, it may be worthwhile briefly reviewing the conceptual basis of the regression analysis and how the basic output parameters of the statistical calculations are used as measures of error and bias. Like Pearson's  $r$ , simple linear regression analysis was not originally conceived as a direct

measure of agreement between two data sets. It is an area of statistics that attempts to predict or estimate an outcome variable from a single predictor variable. Put more specifically, regression analysis is used if the intent is assess to what degree values in one data set (response, dependent, outcome) can be predicted from given values in a another data set (explanatory, independent, predictor). When used for agreement studies, good predictive power of the variables is equated with good agreement. To achieve this, the simple linear regression analysis fits a linear equation to the data that best describes the relationship between the two measurements. In essence the analysis attempts to best summarize the relationship with a straight line equation ( $Y = bX + a$ ). The form of the relationship and predictive power of the variables is then assessed using several output parameters related to this line.

The basic output parameters from regression include the slope and intercept of regression line with confidence intervals, significance test for the slope and the coefficient of determination ( $R^2$  or  $r^2$  if only two data sets). The slope and intercept (with confidence intervals) are output components that describe the linear regression line (line of best fit). These are also known as the regression coefficients. The slope or regression coefficient for the explanatory variable (X values; ‘cause’, ‘predictor’ or ‘independent’) of the straight line of best fit (often called  $b$ ,  $b_1$  or  $m$  in mathematics) essentially indicates how much the average value of outcome variable changes with a unit change in the predictor variable. When applied to agreement studies, this coefficient is used a measure of *proportional bias* between the two data sets (Figure 20); a line that differs from unity ( $Y=X$ ) indicating *proportional bias*. There is often a significance test ( $t$ -statistic test) of whether the slope is significantly different from zero and confidence intervals for the slope. A statistically significant  $P$  value would indicate that there is a relationship between the variables and that predictor variable contributes significantly (at least statistically) to our ability to estimate values of the outcome variable.

The ‘intercept’ or ‘constant’ (also called  $b_0$ ,  $a$  or  $c$ ) is the predicted value of Y when X equals 0. It is also usually accompanied by confidence intervals and a significance test of whether it is significantly different from zero. When simple linear regression is used to quantify agreement, this  $b$  value is interpreted as a measure of the *fixed bias* between measures (Figure 20). As already highlighted in the previous subsection, the coefficient of determination ( $R^2$ ) represents a measure of how well the data fits with the simple

linear regression model. It takes the correlation coefficient a step further to represent the percentage of the independent variation that can be predicted (or accounted for) by the dependent variable. However, as  $R^2$  is derived directly from  $r$ , it is also sensitive to the *random error*, *heteroscedasticity* and range of measures but insensitive to systematic bias (*fixed* and *proportional*). As such, the use of  $R^2$  as a sole indicator of agreement can also be very misleading (Table 14; page 100).



**Figure 20.** Hypothetical scatterplot showing the output components of regression analysis that may be used for the assessment of agreement. Measurement X is the ‘cause’, ‘predictor’, ‘independent’ or ‘explanatory’ variable. Measurement Y is the ‘response’, ‘dependent’, ‘outcome’, ‘effect’ variable. Devised by Rix and Webb (2005), used with permission.

A more direct and specific quantification of the *random error* (in the scale of measurement used) can be obtained using the 95% prediction intervals around a regression line. These intervals represent the area in which you expect 95% of all data points to fall. The prediction intervals are derived from the standard error of the estimate

( $S_\epsilon$  or  $S_y$ ) also known as the standard deviation of the predicted values. In simplistic terms,  $S_\epsilon$  is the spread of the scores about the regression line just as the standard deviation is a spread of scores about the mean and is therefore a direct reflection of the *random error* (Figure 20). The 95% prediction intervals represent the predicted value of  $Y \pm 1.96S_\epsilon$  and as such are the linear regression equivalent to the 95% limits of agreement ( $d \pm 1.96S_d$ ) calculated in the Bland and Altman method of analysis (Bland and Altman, 1986; Bland and Altman, 2003).

***Simple linear regression models and limitations.*** Most of the limitations for the use of simple linear regression for the assessment of agreement are related to the type of regression model used and the inherent mathematical calculations and statistical assumptions. As a background to this discussion, it is worth considering that a general regression (prediction) model for any data set, whatever the nature of the relationship, can be defined using equation 3.1 (Field, 2005). In essence this means that the outcome we are trying to predict for an individual (the  $i$ th subject) can be best predicted by a certain model plus some kind of error. In linear regression, the model that is used is linear in nature; a straight line model that best summarizes the predictive relationship between the data sets. As such, the general model equation becomes equation 3.2. The term  $\epsilon_i$ , also known as the error or residual, represents the difference between the score predicted by the line for subject  $i$  and the score this subject actually obtained. The importance of this term will be discussed shortly. The regression equation (3.3) does not include the error term as it represents the estimated (predicted) value of  $Y$  ( $E[Y]$ ) in a population of  $X$ ,  $Y$  values (Field, 2005).

$$\text{Outcome}_i = (\text{Model}_i) + \text{error}_i \quad 3.1$$

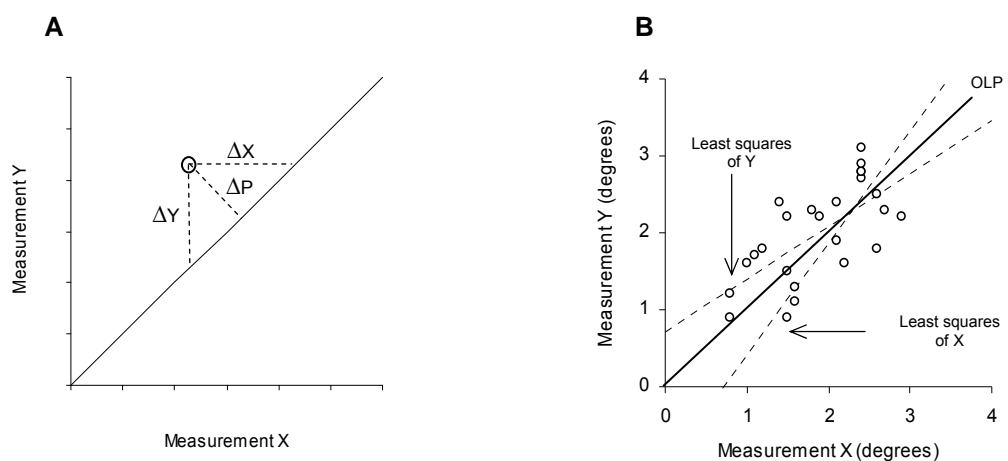
$$Y_i = (bX_i + a) + \epsilon_i \quad 3.2$$

$$E[Y] = bX + a \quad 3.3$$

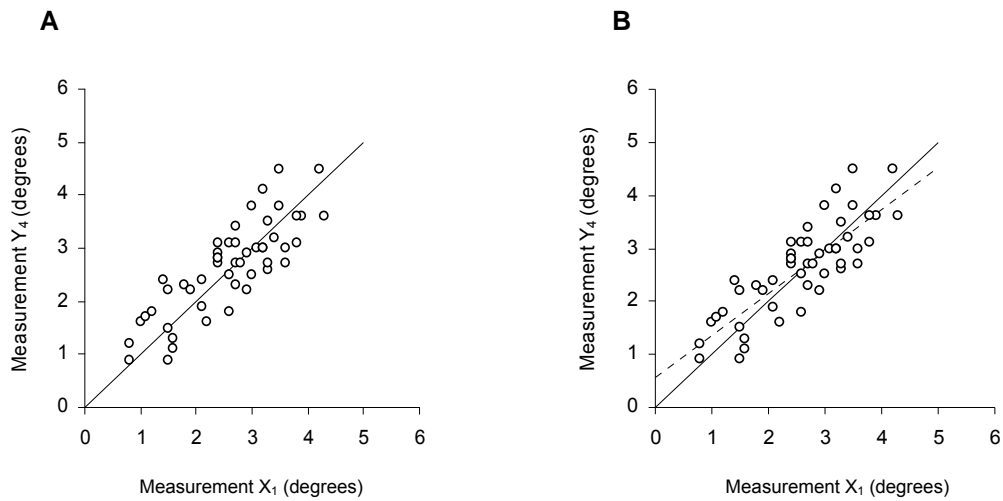
With any data set, several lines could be used to summarize the relationship. For the most accurate conclusions we need to fit a model (straight line) that *best* describes the data with the minimum amount of error ( $\epsilon_i$ ). There are several mathematical ways to do

this. The most familiar form is least squares of Y regression analysis (Model I regression analysis), commonly known as ordinary least squares (OLS) Y regression; this is what is provided by most statistical programs and used in the majority of agreement studies. In establishing a line of best fit, an important assumption of OLS Y regression (Ludbrook, 1997) is that only the values of Y variable (outcome) are attended by error, the values of X variable (predictor) are measured without error or fixed in advance. The error term in the equation Y ( $\epsilon_i$ ) therefore reflects only the vertical deviations of the observed values of Y from the line of best fit ( $\Delta Y$  in Figure 21A) and the OLS line of best fit is determined by calculating the straight line through the data that has the minimum sum of the squared vertical differences.

In the majority of agreement studies both Y and X measurements are attended by error so the OLS assumption is rarely fulfilled. The OLS regression line derived from minimizing the sums of squares of the deviations of Y from the line when both Y and X measurements are attended by error, may result in a reduced slope and an intercept that is raised above zero (Figure 22). The possible implication is that the OLS Y regression coefficients may exaggerate any *proportional* and *fixed bias* or convey the implication that there are such biases between the measurements when none actually exist (Table 14; data sets XY<sub>4</sub>; rows 5, 11 & 14).



**Figure 21.** A) – Residuals types used with variations of linear regression analysis. Solid line – line of best fit;  $\Delta Y$ , vertical distance of x,y point from line;  $\Delta X$ , horizontal distance of x,y point from line;  $\Delta P$  perpendicular distance of x,y point from line. B) – Hypothetical example showing the variation in regression line (dotted line) with Ordinary Least Squares (OLS) of Y and OLS of X regression techniques. Abbreviations; OLP – Ordinary Least Products. Based on Ludbrook (1997)



**Figure 22.** A) – Scatterplot for data set  $X_1Y_4$ ; solid line – line of equality (true relationship measurements; B) Ordinary Least Squares (OLS) of Y regression analysis for data set  $X_1Y_4$  showing the attenuation of regression line with this technique (dotted line).

As with correlation coefficients, when the *random error* is small relative to the range of values (data set XY), the estimates of bias may be relatively unaffected (Table 14; row 1). However, the greater the *random error* is relative to the range (data set  $X_1Y_4$ ), the more the bias estimates are exaggerated with the OLS model (Table 14; rows 5, 11 & 12). The effects are greatest when measurements are made over narrow range of values located some distance from zero (Table 14; row 12) (Westgard and Hunt, 1973; Ludbrook, 1997). It should be highlighted that although the slope and intercept can be affected by the range of measurements, the standard error of the estimate ( $S_e$  or  $S_y$ ) remains relatively constant with a change in the range of the data thereby providing a stable estimate of *random error* under these conditions (Table 14; data set XY; rows 1, 9 & 10; data set  $X_1Y_4$ ; rows 5, 11 & 12). The  $S_e$  also remains relatively unaffected by a *fixed bias* (Table 14; data sets XY &  $XY_1$ ; rows 1 & 2; data set  $X_1Y_4$ ; rows 5 & 6). However, the introduction of a *proportional bias* does lead to an increase in the  $S_e$  with no real change in the *random error* (Table 14 data sets XY &  $XY_2$ ; rows 1 & 3; data set  $X_1Y_4$ ; rows 5 & 7). From this it can be seen that with OLS regression, the presence of a *proportional bias* may lead to an exaggerated estimate of *random error* and thus may be misleading.

**Table 14.** Effects of various biases and errors on statistical results from Ordinary Least Squares (OLS) Simple Linear Regression analysis. Data is artificially generated with known errors and biases introduced

Data sets	n	$r^2$	Regression coefficients		S $\epsilon$	P-Values	
			Slope - $b_1$ [95% CI]	Y-Intercept - $b_0$ [95% CI] (degrees)		Slope	Intercept
XY	50	0.96	1.0 [0.95 → 1.08]	0.0 [-0.21 → 0.14]	0.19	< 0.001	0.68
XY <sub>1</sub> (fixed) <sup>†</sup>	50	0.96	1.0 [0.95 → 1.08]	1.0 [0.79 → 1.14]	0.19	< 0.001	< 0.001
XY <sub>2</sub> (proportional) <sup>‡</sup>	50	0.96	1.5 [1.42 → 1.61]	0.0 [-0.29 → 0.25]	0.29	< 0.001	0.89
XY <sub>3</sub> (fixed and proportional)	50	0.96	1.5 [1.42 → 1.61]	1.0 [0.72 → 1.25]	0.29	< 0.001	< 0.001
X <sub>1</sub> Y <sub>4</sub> ( $\uparrow$ RE)	50	0.68	0.8 [0.61 → 0.93]	0.6 [0.22 → 1.07]	0.50	< 0.001	< 0.01
X <sub>1</sub> Y <sub>4</sub> (+ fixed) <sup>†</sup>	50	0.68	0.8 [0.62 → 0.93]	1.6 [1.22 → 2.07]	0.50	< 0.001	< 0.01
X <sub>1</sub> Y <sub>4</sub> (+ proportional) <sup>‡</sup>	50	0.68	1.2 [0.93 → 1.34]	1.0 [0.37 → 1.63]	0.75	< 0.001	< 0.01
XY <sub>5</sub> (heteroscedasticity)	50	0.55	0.8 [0.60 → 1.02]	0.5 [-0.06 → 1.09]	0.63	< 0.001	0.08
XY (range 0 to 2.5°)	25	0.88	1.0 [0.87 → 1.20]	-0.1 [-0.42 → 0.25]	0.20	< 0.001	0.60
XY (range 2.6 to 5.0°)	25	0.90	1.0 [0.81 → 1.09]	0.2 [-0.30 → 0.63]	0.19	< 0.001	0.47
X <sub>1</sub> Y <sub>4</sub> (range 0 to 2.5°)	25	0.45	0.7 [0.36 → 1.01]	0.7 [0.07 → 1.37]	0.48	< 0.001	< 0.05
X <sub>1</sub> Y <sub>4</sub> (range 2.6 to 5.0°)	25	0.21	0.5 [0.08 → 1.11]	1.5 [-0.09 → 3.01]	0.52	< 0.01	0.06
XY (n=25)	25	0.96	1.0 [0.90 → 1.10]	0.0 [-0.23 → 0.28]	0.20	< 0.001	0.83
X <sub>1</sub> Y <sub>4</sub> (n=25)	25	0.73	0.9 [0.63 → 1.00]	0.3 [-0.31 → 0.92]	0.55	< 0.001	0.31

*Abbreviations:*  $\uparrow$ RE – increased random error compared with XY data set; n – sample size; S $\epsilon$  – standard error of the estimate (standard deviation of predicted values)

<sup>†</sup> Fixed (additive) bias of +1.0 degrees to Y data

<sup>‡</sup> Proportional (multiplicative) bias of 1.5 degrees to Y data

A further limitation of OLS regression analysis arises when you consider that the line resulting from minimizing the sums of the squares of the deviations of the Y values from the line (least squares of Y) and that resulting from minimizing the sums of the squares of the deviations of the X values from the line (least squares of X) are distinctly and sometimes markedly different (Figure 21B). In most agreement studies, it is not possible to decide which measurement should be regarded as the dependent and which the independent variable. In this context neither line could be considered the most appropriate or a good illustration of the interrelation of Y and X.

**Model II linear regression.** There is an alternative the OLS regression model (Model I). It is the Model II regression analysis. This analysis method is designed for cases in which both X and Y values are attended by error and especially when it is impossible to

decide which should be regarded as dependent and independent variables. In Model II regression, the deviations of both Y and X values from the fitted line are minimized. As a result there is only one line to describe the relationship in Model II compared with two for Model I (OLS). This can be regarded as a line of symmetry between the two OLS best fit lines (Figure 21B; page 98).

There are several techniques for taking into account the errors attached to both X and Y in Model II regression analysis but only two appear to be in common use. These are *major axis regression* (principal component regression or perpendicular distance method) and *ordinary least products* (OLP) regression (reduced major axis regression, standardized principal component regression or geometric mean regression). The major axis regression analysis fits a straight line to the data by minimizing the sum of the squared perpendicular ( $\Delta P$  in Figure 21A) distances of the X,Y values from the line -  $\sum(\Delta P)^2$ . In OLP regression the line is derived by minimizing the sum of the products of the vertical ( $\Delta Y$  in Figure 21A) and horizontal ( $\Delta X$  in Figure 21A) distances of the X,Y values from the line -  $\sum(\Delta Y)(\Delta X)$ .

Except that both X and Y can be attended by error, Model II regression analysis depends on a similar set of assumptions to OLS (Model I). In addition, the output parameters for Model II regression analysis are in the same form as for Model I (OLS) with regression coefficients, a coefficient of determination, standard error of the estimate, confidence intervals and significance tests. However, unlike Model I regression (OLS), the output parameters for Model II analyses are not affected by the range of data. Although Model II analysis is more suitable for agreement studies, the analysis method is not easy to find within common statistical software packages and is generally poorly supported. With OLP regression, the regression coefficients can be calculated by hand. Calculating the 95% confidence intervals can be more problematic. A detailed description of how to perform these calculations is provided by Ludbrook (1997)

***Limitations of Model II regression.*** Although Model II linear regression provides an appropriate method of analysis when both measurements are attended by error, there are still some difficulties and limitations with its use. Firstly, as with several of the other approaches, the output parameters from both Model I and II linear regression analyses are affected by sample size. Secondly, an important assumption of both OLS and OLP



regression is that the variance in differences between measurements is the same across all the measurements (*homoscedasticity*). However, as highlighted earlier, biological data often exhibit *heteroscedasticity* (Figure 13C). In this instance, the 95% prediction intervals will be too wide for the measurements in the low range and too narrow for the higher values. As such, an erroneous estimate of *random error* will be produced. As with the Bland and Altman method, there are two ways of coping with this relationship. The simplest approach is to logarithmically transform both the X and Y values before undertaking the regression analysis. However, it leads to difficulties in making inferences about the presence of *proportional* or *fixed bias*. An alternative approach, which avoids these problems, is to weight the values of X and Y using a weighted least products regression (WLP).

### **Summary: Simple linear regression as a measure of agreement**

To summarize then, although originally conceived as a predictive power analysis tool, the use of an appropriate linear regression model (Model II - OLP) does provide a very useful method for assessing agreement between measurements of an interval nature. It is sensitive to *random error* and *systematic bias* and provides a specific quantification of each in the scale of measurements used for the tests. It also provides a method of analysis when the assumption of homoscedasticity is violated. Although, all types of linear regression are affected by the sample size, the most misleading results arise from the inappropriate use of OLS regression methods. This approach may lead to erroneous inferences regarding the systematic biases present between measures particularly when the *random error* is large and the range of data is small and further from zero.

### **3.3.4 Paired t-test**

The paired *t*-test (also referred to as the Student's or dependent *t*-test) compares two paired groups to make inferences about the size of the average difference between the paired measurements. When this analysis of difference is used as a complete measure of *reliability*, the investigator concludes that an acceptable level of agreement has been demonstrated if there is no significant difference between the data sets. A significance difference would indicate poor agreement. When used as a *reliability* index in this way, the *t*-test can be quite misleading and has little value; there are a large number of instances when the level of agreement can be clearly unacceptable for a given use but no

significant difference is demonstrated and vice versa. In addition, the output parameters from a paired  $t$ -test are sensitive to more than one error or bias types and therefore do not provide a specific quantification of each (Table 11). As a result, misleading inferences may be drawn on the types and size of and biases and errors present without a closer visual examination of the data and/or complementary analysis of the data. To illustrate the problems and the inappropriate uses of the  $t$ -test outcomes in method agreement and precision assessment, some numerical examples are given. First though, it is helpful to look at the  $t$ -test rationale and design.

**Paired  $t$ -test model & calculation.** The paired  $t$ -test is essentially a test that the average of the paired differences is zero (null hypothesis or  $H_0$ ). This is achieved by calculating a  $t$ -value (statistic) using the following equation (3.4). This equation compares the mean of the paired differences between our data sets ( $D = x_{ij} - y_{ij}$ ) with the difference that we would expect to find if the null hypothesis were true ( $\mu_D$ ) and taking into account the standard error of the differences – SE(d) ( $S_D/\sqrt{N}$ ). As the null hypothesis is that there is no difference then  $\mu_D$  is 0. Hence the equation becomes 3.5

$$t = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}} \quad 3.4$$

$$t = \frac{\bar{D}}{S_D / \sqrt{N}} \quad 3.5$$

$D$  – mean difference between samples;  
 $\mu_D$  – mean expected to find between population means  
 $S_D/\sqrt{N}$  – standard error of differences

This  $t$ -statistic is then interpreted by comparing the calculated  $t$ -value with the critical value ( $t_{crit}$ ). This critical value can be found in statistics table at a given degrees of freedom (df) and probability level (n-1 df;  $\alpha = 0.05$ ). A difference between the means (average of the paired differences to zero) is significant if the calculated  $t$ -value is greater than the value given in this table (at the given probability level). The  $t$ -test is often interpreted using the confidence intervals for the difference (95% CI for d) and the P-value. The 95% CI for d is the most valuable calculation. If this does not cross zero, then difference is likely to be statistically significant. The confidence interval for the average of these paired differences is particularly important when judging the scientific

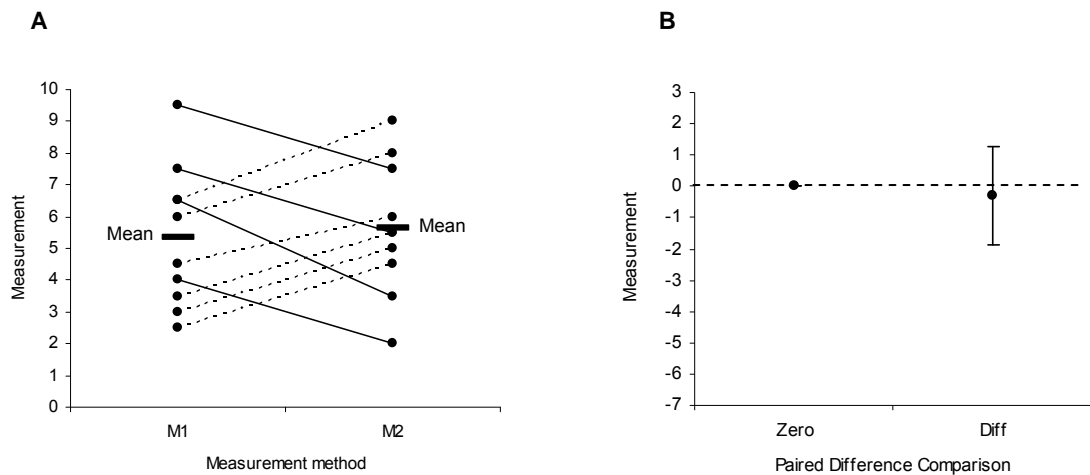
or clinical significance. The  $P$ -value is relatively meaningless in isolation but a value less than the set  $\alpha$  level indicates that the  $t$ -value exceeds the  $t_{crit}$  and in most cases that the 95% CI for  $d$  does not cross zero. In addition to the mathematical view of the  $t$ -test (mean of paired diffs/standard error of diffs), conceptually the paired  $t$ -statistic could be viewed as a ratio of the systematic variation (or bias) to unsystematic variation (or *random error*). Others view it as a type of ‘signal to noise ratio’ (Weir, 2005).

$$\frac{\text{mean of paired difference}}{\text{standard error of paired difference}} \quad \frac{\text{systematic variance (bias)}}{\text{unsystematic variance (random error)}}$$

***The Paired t-test: The effect of error/bias types.*** From these equations it is clear that the detection of a significant difference is dependent on the amount of random variation between the data sets. This mathematical relationship can lead to erroneous judgements when assessing agreement using the results of a  $t$ -test, particularly when the  $t$ -value is the only parameter considered (Westgard and Hunt, 1973). When the  $t$ -test is used as a reliability index, the absence of significant difference between the data sets infers that an acceptable level of agreement has been demonstrated and vice versa. However, poor agreement can exist between data sets when the result of a  $t$ -test shows no significant difference (small  $t$ -value). Conversely, acceptable agreement can exist when a statistically significant difference (large  $t$ -value) is demonstrated using a  $t$ -test. Some numerical examples using artificially generated data may help demonstrate the problem better.

In the first example, applying a two-tailed paired  $t$ -test to the data in Figure 23, infers that there is no statistically significant difference between the data sets ( $d = -0.30$ ; 95% CI = -1.89 to 1.29;  $t = -0.43$ ;  $df = 9$ ;  $P = 0.68$ ). With the use of these  $t$ -test parameters in isolation, agreement would be concluded as acceptable. However, the agreement may not be considered so acceptable when the within subject variation (agreement between data pairs) is viewed more closely. From Figure 23A, widely discrepant values between the data sets can be seen; this represents a relatively large amount of *random error* and is also reflected in the standard deviation of the differences ( $S_d = \pm 2.23$ ). However, this random variation is relatively symmetrical around zero; some scores increasing and

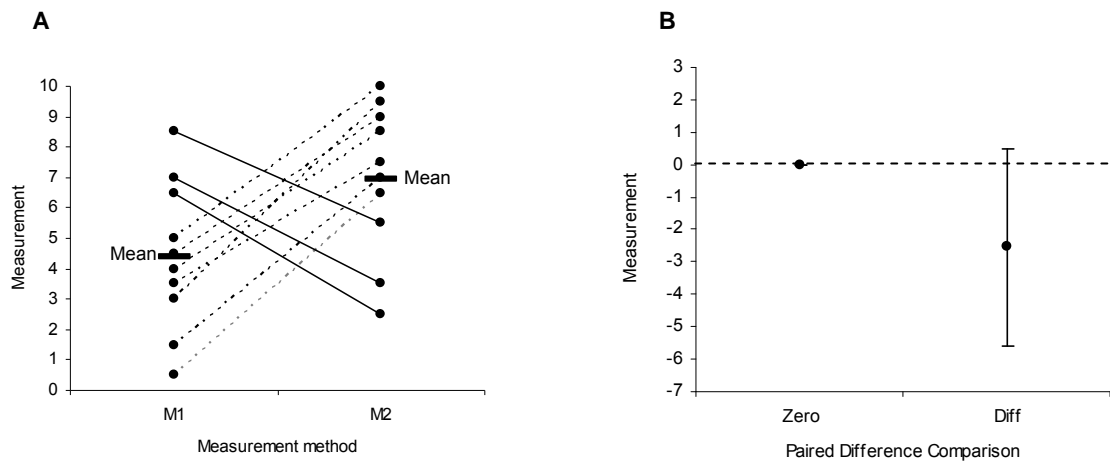
others decreasing resulting in a mean paired difference close to zero. As the ratio of *systematic bias* (mean difference) to *random error* ( $s_d$ ) is quite small, a relatively small and statistically insignificant  $t$ -value is produced. In other words, the *random error*, and hence potentially poor agreement, is hidden by the  $t$ -test results. Put another way, from the equations earlier, as the  $S_d$  (*random error*) goes up with respect to the systematic bias (mean of paired differences), the  $t$ -value gets smaller implying an increasing level of agreement; quite the opposite of what is actually happening. In some cases the *random error* can be very large indeed, potentially making the test characteristics unacceptable. However, this judgment is ultimately dependent on the intended purpose of the test.



**Figure 23.** A) - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set. B) – Graph representing the mean of the paired differences compared with zero (no difference). The error bar represents the 95% CI for the difference. Devised by Webb and Rix (2005), used with permission.

The second example demonstrates a further possible compromising effect of large amounts of *random error* on the results of the paired  $t$ -test. When a two-tailed paired  $t$ -test is performed on the data shown in Figure 24, the mean of the paired differences is not quite statistically significant ( $d = -2.55$ ; 95% CI =  $-5.59$  to  $0.49$ ;  $t = -1.90$ ;  $df = 9$ ;  $P = 0.09$ ). As with the former example, this would lead to a conclusion that there is acceptable agreement between the measurements. However, when the differences

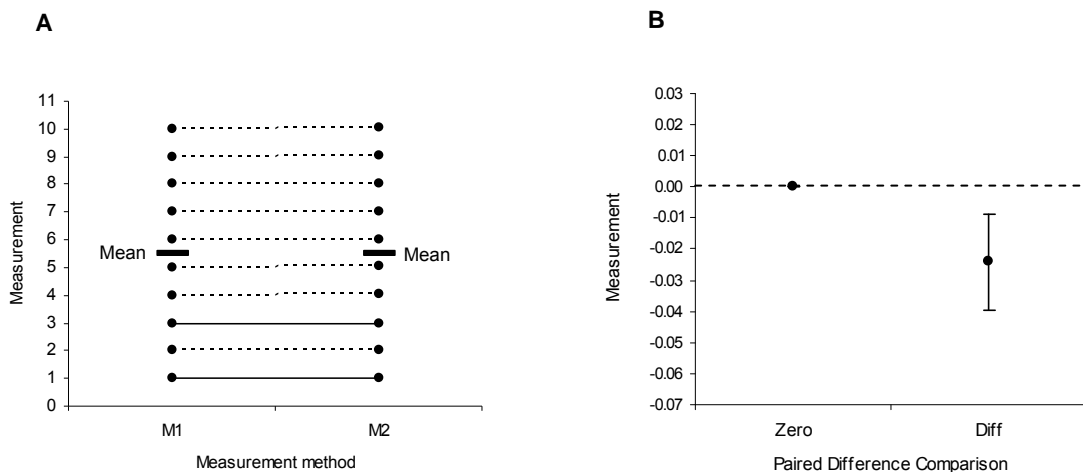
between the measurements are examined more closely, the level of agreement could be considered relatively poor. There is clearly a relatively large *random error* present ( $S_d = 4.25$ ) but with more values increasing than decreasing (Figure 24); resulting in a relatively large mean difference ( $d = -2.55$ ). This suggestion of a *systematic bias* between the measurements, does not quite reach a statistically significant level mainly due to the compromising effect of the relatively larger *random error*. Although 3 of the 10 measurements do vary in the opposite direction, the systematic error seen with the other measurements may be clinically meaningful. This example highlights that a significant systematic bias will be less likely to be detected if it is accompanied by large amounts of *random error*, particularly when the sample size is low. It could be argued that increasing the sample size would increase the power to detect a significance difference but as highlighted a little later, the relationship of sample size to levels of agreement can also be misleading.



**Figure 24.** A - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set. B – Graph representing the mean of the paired differences compared with zero (no difference). The error bar represents the 95% CI for the difference. DeVos and Rix (2005), used with permission.

The next example highlights the potential misleading nature of the paired *t*-test results for assessing agreement when both the *random error* ( $S_d$ ) and systematic bias (mean paired difference,  $d$ ) are both small. The results of a two-tailed paired *t*-test performed

on the data shown in Figure 25, show a statistically significant difference ( $d = -0.02$ ; 95% CI = -0.04 to -0.008;  $t = -3.50$ ;  $df = 9$ ;  $P < 0.01$ ) thereby implying an unacceptable level of agreement (note difference in Y-axis scales for Figure 25A & B). However, when the differences between the measurements are examined more closely (Figure 25A), the level of agreement could be considered relatively good with a very small random error ( $S_d = 0.02$ ). This example further highlights the compromising effect of the  $S_d$  and shows that even small differences between data sets, as long as they are in the same direction (consistent across all subjects) can result in a statistically significant effect using a paired  $t$ -test. The use of these  $t$ -test results alone may lead to the rejection of a testing procedure or measurement method when the actual size and nature of errors is quite acceptable for the intended analytical goals.



**Figure 25.** A - Graph of paired data showing individual differences between measurements M1 & M2 and the mean for each data set (dotted & solid lines represent difference in a -ve or +ve direction). B – Graph representing the mean of the paired differences compared with zero (no difference). The error bar represents the 95% CI for the difference. Devised by Webb and Rix (2005), used with permission.

***The Paired t-test and identification of error/bias types*** As highlighted at the start of this section, the output parameters from a paired  $t$ -test are sensitive to more than one error or bias types (Table 11; page 75). A mean paired difference between data sets is a reflection of a *systematic bias*. However, this bias may be *fixed* or *proportional* in nature. If data set  $XY_1$  (*fixed bias*) (Appendix II) is compared using a two-tailed paired

*t*-test, a statistically significant difference is demonstrated ( $d = -0.99$ ; 95% CI = -1.05 to -0.94;  $t = -36.8$ ;  $df = 49$ ;  $P < 0.0001$ ). When a two-tailed paired *t*-test is performed on the data set exhibiting a *proportional bias* ( $XY_2$ ), a similar statistically significant difference is also demonstrated ( $d = -1.31$ ; 95% CI = -1.47 to -1.17;  $t = -17.5$ ;  $df = 49$ ;  $P < 0.0001$ ). This clearly demonstrates that the bias type cannot be inferred from *d* as an output parameter without a closer visual examination of the data and/or complementary more specific analysis of the data. It is also interesting to note that agreement for these data sets would be considered acceptable when Pearson's *r* is used as the agreement index (Table 12) and unacceptable using a paired *t*-test. Further misleading inferences can also arise from  $S_d$  (standard deviation of the differences). This paired *t*-test output parameter provides an estimate of the *random error*, but is also sensitive to *proportional error* and *heteroscedasticity* in data sets (Table 11). The effect of *proportional bias* is nicely highlighted using data sets  $XY$  (*random error*) and  $XY_2$  (*random error* and *proportional bias*). When only *random error* is present ( $XY$ ), the  $S_d$  is 0.19. However, when an additional *proportional bias* is introduced ( $XY_2$ ),  $S_d$  becomes 0.53. Thus, it can be seen that the presence of *proportional bias* may lead to an over-estimate of *random error*. When *heteroscedasticity* is present, the  $S_d$  provides only a crude estimate of the *random error*. In this instance, the  $S_d$  will be too wide for the measurements in the low range and too narrow for the higher values. Finally, for given error levels, the *t*-value will vary depending on the sample size (*n*). As *n* increases, the *t*-value will increase reflecting a decrease in the  $SE(d)$  in the *t*-test formula. The consequence of this may be the acceptability of measurement agreement being related to sample size in the study; the inference of acceptable agreement with low sample sizes and unacceptable agreement with larger sample sizes.

### **Summary: Paired *t*-test as a measure of agreement**

To summarize, the statistical parameters used in a paired *t*-test make this test sensitive to systematic bias (*fixed* and *proportional*) and *random error*. However, the nature of the *t*-test formula dictates that when utilised for making inferences regarding overall agreement between measurements, the hypothesis test results of a paired *t*-test can be misleading; particularly when the *t*-value is viewed alone. A failure to show a statistically significant difference between the data sets is not always the same as acceptable agreement. Conversely, demonstrating a statistically significant difference, does not always equate with unacceptable agreement. The misleading results arise

mainly because of the compromising effects of the  $S_d$  in the  $t$ -test calculations. Finally, analysis by  $t$ -test can provide specific estimates of *random error* and *fixed bias*, but only when *proportional error* and *heteroscedasticity* are absent.

### 3.3.5 Bland & Altman's 95% Limits of Agreement

This method of analysis, also commonly referred to as the *method of differences* or *limits of agreement* (Hopkins, 2000; Ludbrook, 2002; White and Van Den Broek, 2004), was originally proposed in 1983 by Altman and Bland (Altman and Bland, 1983). It was developed mainly to overcome several of the limitations of the other methods of measuring agreement; particularly those associated with correlation coefficients and OLS linear regression. However, this approach really only made an impact when it was proposed again in the Lancet (Bland and Altman, 1986). As the order of the authors' names was reversed for this publication, the approach subsequently became known as the Bland and Altman method of agreement. The Bland and Altman method has gained considerable popularity over the years and has most often been described within the framework of method agreement and repeatability studies as originally proposed. However, it is equally suitable for studies examining test-retest and inter-rater precision.

***Bland & Altman model, analysis procedure and calculation.*** Before discussing the benefits and limitations of the Bland and Altman method, it might be helpful to review the analysis procedure. A detailed description of steps in analysis is provided in the Chapter 4 (methods section) and Appendix III. In brief, the Bland & Altman analysis steps and basic outcome measures are the following:

- Distribution plots (Figure 26)(difference of two measurements against the their mean).
- Mean difference between methods ( $d$  – *fixed bias*) with CI's.
- Limits of agreement (LoA – *random error*) with CI's.
- Modified (regression adjusted)  $d$  & LoA's (*proportional bias & heteroscedasticity*).

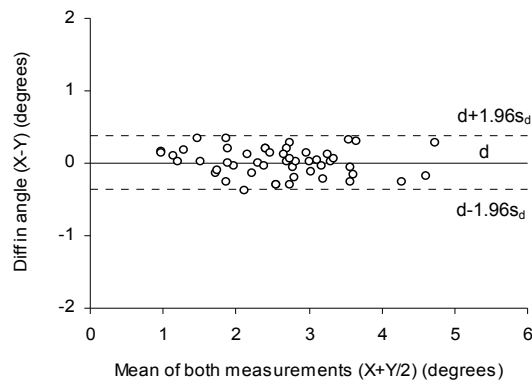
Although the authors have persistently reiterated and developed their approach over the years, often with worked examples, guidance regarding the application of the Bland and



Altman method is sometimes incomplete and occasionally unclear. However, overall the method does provide a relatively simple and easily understood approach using simple graphical techniques and elementary statistical calculations. The authors suggest that these can be performed by anyone with only basic statistical knowledge (Bland and Altman, 2003). This approach provides an estimation of the bias and error between two measurement methods and determines the limits within which most of the differences will lie; the 95% limits of agreement (Bland and Altman, 1986).

The first stage in the analysis requires a calculation of the difference in measurement values obtained by the two measurement methods (or test-retest intervals etc.) and the mean of these paired values. Before calculating the limits of agreement, two assumptions are checked statistically and graphically (distribution plot-differences between paired values against the mean of the paired values). The differences calculated are first checked for a normal distribution. Assuming a normal distribution, a distribution plot is then constructed (Figure 26) so that the  $s_d$  of the differences can be visualised; if no *proportional bias* or *heteroscedasticity* is present, the  $s_d$  should be constant across the range of mean scores (X-axis). If the  $s_d$  is constant, then the mean difference of the two measurement values (*fixed bias*) limits and the limits of agreement (LoA) are calculated (*random error*). The LoA are calculated as  $d \pm 1.96s_d$ . Both  $d$  and  $s_d$  can be represented on the distribution plot (Figure 26). The mean difference ( $d$ ) and LoA can be calculated in the presence of a *proportional bias* or *heteroscedasticity* bias using a modified approach. This involves regression adjustment of  $d$  and the LoA's (see Chapter 4 – methods section & Appendix III)

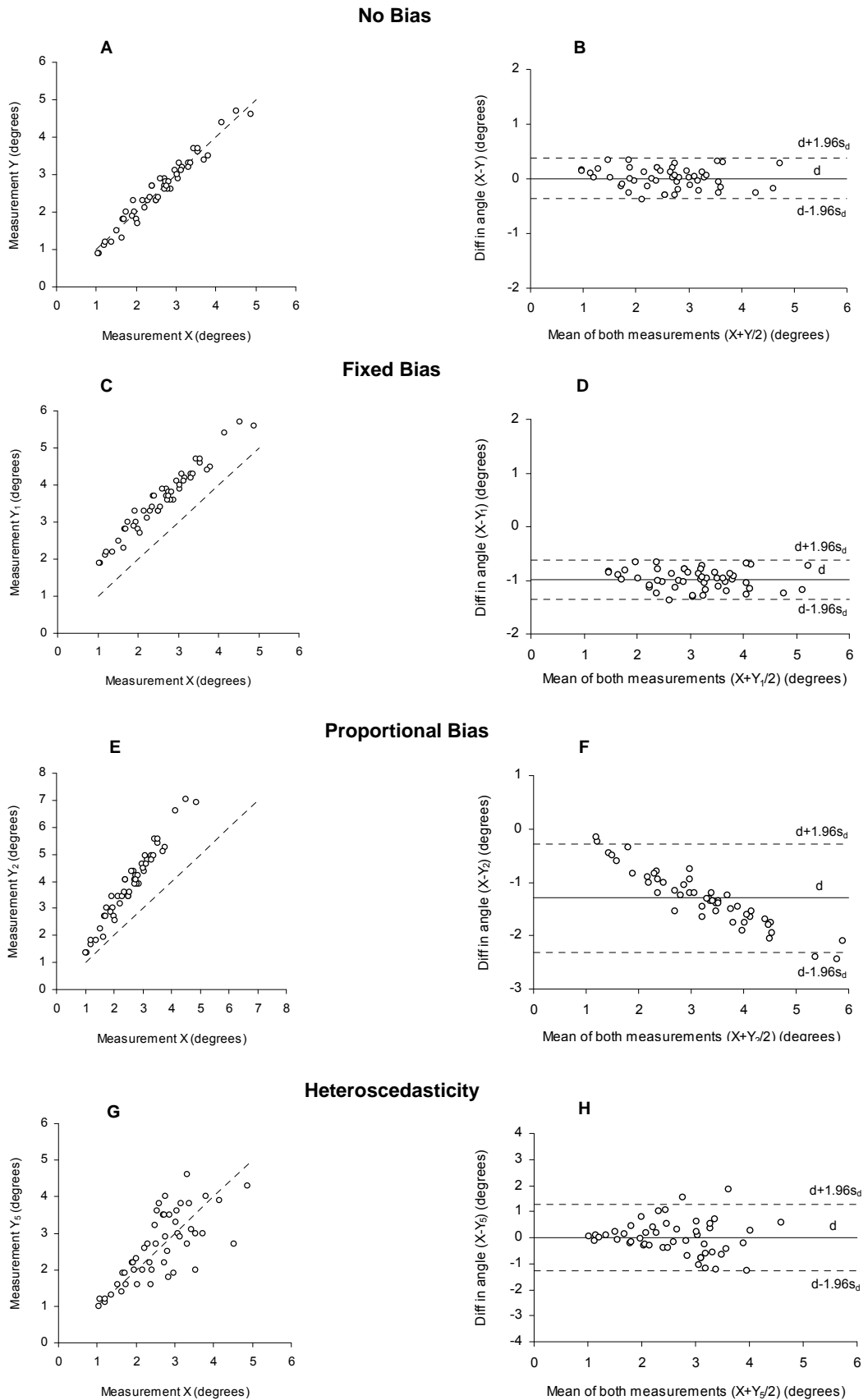
***The Bland Altman approach and identification of error/bias types.*** A useful advantage of the Bland and Altman method (as with linear regression), compared with correlation coefficients in particular, is by establishing the limits of agreement in the scale of measurement it provides easily interpreted quantification in a meaningful form. By doing this, it allows the researcher or method user to make the most important clinical judgement of acceptability for a given purpose; the question essentially being “are the LoA's small enough for us to say that the methods or raters agree sufficiently for the clinical purpose of the test?” (Bland and Altman, 2003).



**Figure 26.** Bland and Altman distribution plot. The dotted lines represent the limits of agreement.

As with the scatterplots shown earlier the use of distribution plots in the Bland and Altman method, enables the pattern of discord to be visualised (e.g. *random error* and *systematic bias*) thus providing a powerful graphical representation of agreement. The distribution plot is easy to interpret allowing a clear visualization of the size and range of differences, the range of variation between subjects, any bias (*proportional* or *fixed*), any extreme or outlying observations, and the relationship between the variance in the differences with the size of mean. The investigator can also check the assumption that the measurements are related in a linear fashion (Lee et al., 1989). A comparison of these patterns between a scatterplot and distribution plot is shown in Figure 27.

It has been argued that traditional scatterplots show the error and bias types in a manner more directly related to a line of equality and in a form that biomedical investigators are used to viewing (Ludbrook, 2002). However, the use of a scatterplot may not allow adequate visualisation of a *proportional bias* particularly if the *random error* is small (tight clustering around the line of equality) and the range of variation measurements is large (Bland and Altman, 1999). A distribution plot (Figure 27F) will often reveal this mainly because of the relative scale of the differences axis (Y-axis) compared with the mean of the two measurements (X-axis). Erroneous inferences arising from either of these helpful but rather crude approaches to examining the data, are avoided by using OLP regression with data presented using a traditional scatterplot and using correlation coefficients for the Bland and Altman approach (as described in Chapter 4 -methods section & Appendix III).



**Figure 27.** Graphical comparison of error and bias types using scatterplots (A,C,E,G) and Bland and Altman distribution plots (B,D,F,H). For scatterplots, the dotted line represents the line of equality. For the distribution plots, the dotted lines represent the limits of agreement.

A graphical representation of agreement and testing for *proportional bias* and/or *heteroscedasticity* (using correlation coefficient or regression – see Chapter 4 methods section & Appendix III) is an extremely important step in assessing the degree of disagreement. Without this, the mean difference and LoA's may well be misleading (Ludbrook, 1997; Bland and Altman, 1999; Ludbrook, 2002). Firstly, within the Bland and Altman method, the *fixed* and *proportional biases* are not seen as independent. It is quite wrong to automatically attribute any mean difference between the measures (M1-M2) and zero to a *fixed bias* when it can equally be accounted for by a *proportional bias* (Figure 27 D & F) or a combination of *proportional* and *fixed bias* both acting in the same direction (Table 15; data set XY<sub>3</sub>; row 4). Conversely, if no mean difference were found between measures (M1-M2) and zero, this could result from there being a *proportional bias* in one direction and a *fixed bias* in the opposite (Ludbrook, 1997; Bland and Altman, 1999; Ludbrook, 2002). In addition, the LoA will be misleadingly wide when a *proportional bias* is present and represent only a crude estimate of the true agreement (Figure 27F and Table 15; data set XY<sub>2</sub>; row 3).

**Table 15.** Effects of various biases and errors on statistical results from Bland and Altman method of analysis. Data is artificially generated with known errors and biases introduced

Data sets	n	Bland and Altman parameters					
		d	1.96s <sub>d</sub>	95% CI for d	95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
XY	50	0.01	0.37	-0.04 → 0.07	-0.36 → 0.38	-0.44 → -0.27	0.29 → 0.47
XY <sub>1</sub> (fixed) <sup>†</sup>	50	-0.99	0.37	-1.04 → 0.93	-1.36 → -0.62	-1.44 → -1.27	-0.71 → -0.53
XY <sub>2</sub> (proportional) <sup>‡</sup>	50	-1.28	1.04	-1.43 → -1.13	-2.32 → -0.25	-2.57 → -2.07	-0.50 → 0.00
XY <sub>3</sub> (fixed and prop)	50	-2.28	1.04	-2.43 → -2.13	-3.32 → -1.25	-3.57 → -3.07	-1.50 → 1.00
X <sub>1</sub> Y <sub>4</sub> (↑RE)	50	-0.06	1.06	-0.21 → 0.09	-1.11 → 0.99	-1.36 → -0.86	0.74 → 1.25
XY <sub>5</sub> (heteroscedasticity)	50	0.00	1.27	-0.19 → 0.18	-1.28 → 1.27	-1.59 → -0.97	0.96 → 1.58
XY (range 0 to 2.5°)	25	0.02	0.37	-0.06 → 0.10	-0.36 → 0.40	-0.49 → -0.23	0.27 → 0.53
XY (range 2.6 to 5.0°)	25	-0.01	0.37	-0.09 → 0.07	-0.38 → 0.36	-0.51 → -0.26	0.23 → 0.49
X <sub>1</sub> Y <sub>4</sub> (range 0 to 2.5°)	25	-0.15	1.04	-0.37 → 0.06	-1.18 → 0.88	-1.53 → -0.83	0.53 → 1.23
X <sub>1</sub> Y <sub>4</sub> (range 2.6 to 5.0°)	25	0.04	1.06	-0.19 → 0.26	-1.03 → 1.10	-1.39 → -0.66	0.74 → 1.46
X <sub>1</sub> Y (half group)	25	0.00	0.37	-0.07 → 0.08	-0.37 → 0.38	-0.50 → -0.25	0.26 → 0.51
X <sub>1</sub> Y <sub>4</sub> (half group)	25	0.06	1.06	-0.17 → 0.29	-1.03 → 1.16	-1.40 → -0.65	0.78 → 1.53

*Abbreviations:* n – sample size; d – difference between measures (X – Y); S<sub>d</sub> – standard deviation of difference; LoA – limits of agreement; ↑RE – increased random error compared with XY data set

<sup>†</sup> Fixed (additive) bias of +1.0 degrees

<sup>‡</sup> Proportional (multiplicative) bias of 1.5 degrees

The LoA's can also be misleading when there is *heteroscedasticity* in the data (Figure 27H and Table 15; data set XY<sub>5</sub>; row 6). In this instance, the LoA will be too wide for the measurements in the low range and too narrow for the higher values. Ludbrook (2002) suggests that if the method of OLP regression analysis is used, the confounding effect of *proportional bias* on the estimation of *fixed bias* is directly avoided.

However, as mentioned earlier, when a relationship exists between the magnitude of the measurement and difference (*proportional bias* or *heteroscedasticity*) the mean difference (M1-M2) compared with zero (*fixed bias*) and LoA's can be correctly established using a relatively simple linear regression (OLS) approach (Bland and Altman, 1999) (Chapter 4 – methods section & Appendix III). Bland and Altman (1986) originally suggested a log transformation of the original data when there are non-uniform differences; this was reinforced by Chinn (Chinn, 1990). However, as this transformation presents the output in proportions rather than the original units, it creates new problems with interpretation (Bland and Altman, 1986; Ludbrook, 1997; Bland and Altman, 1999).

An important property of the Bland and Altman 95% Limits of Agreement approach is that it is unaffected by the range of measurements (Table 15; rows 1, 5, 7-12); a useful advantage over correlation coefficients and linear regression when comparing between different populations and measurement tools. However, when a relationship exists between the magnitude of the measurement and difference, the LoA will be affected by the range of measures unless the data is log transformed or analysed using the regression method (Table 15; data sets XY<sub>2</sub> & XY<sub>5</sub>; rows 3 & 6).

**Reporting Bland & Altman outcome results.** When reporting the results of a Bland and Altman analysis, investigators often produce the LoA in isolation (e.g., -2.5 to +3.60 degrees). Expressed this way, the LoA are actually a measure of total error. As highlighted above, this may include *proportional bias* and *heteroscedasticity* as sources of 'error' if the data has not been properly examined before proceeding to the LoA calculation. If these relationships are not present, the total error constitutes only the *random error* and *fixed bias* together. If the LoA's are symmetrical around zero, then there is a rationale for expressing the results in an even shorter form as  $\pm 1.96S_d$  (e.g.  $\pm 3.05$ ). However, if there is no significant systematic bias (identified by a paired *t*-test),

but the LoA's are asymmetrical around zero, the results should not be expressed as a  $\pm$  value. One of the discussed drawbacks of the paired *t*-test was that clinically significant bias may not be detected if it is accompanied by large *random error*. When the LoA's are not symmetrical about zero it is generally more informative if the *fixed bias* and *random error* are cited separately even when the difference is not statistically significant, e.g.,  $0.55 \pm 3.05$  degrees as opposed to LoA's of -2.5 to +3.60. This is particularly important for measurements of human performance which are likely to be affected by learning or fatigue.

***Limitations of Bland and Altman approach to assessing agreement.*** The previous points of discussion do not represent limitations of the Bland and Altman method. Rather, they serve to highlight the possibility of misleading conclusions when Bland and Altman method is not appropriately applied. Their analyses dictate that the LoA and CI's are affected by the sample size; unacceptably large LoA's and CI's may occur with small sample sizes. Sample sizes of at least 50 individuals in a study have been recommended in order for the sample LoA to be a precise estimate of the population LoA. The Bland and Altman approach can also only be used for comparison of two measures. To the author's knowledge, there is no modification available for when more than two measures are compared at once.

### **Summary: Bland & Altman method as measure of agreement**

To summarize then, like Model II linear regression, the Bland and Altman approach provides a very useful method for assessing agreement between two measurements of an interval nature. The outcome measurements are provided in the units measured and the nature of any biases or errors present can easily be visualised using a distribution plot. These plots have an advantage over traditional scatterplots in that a small *proportional bias* can be visualised particularly if the *random error* is relatively small. The Bland & Altman outcome measures provide a measure of the *fixed bias* (*d*) & *random error* ( $\pm 1.96s_d$ ). Misleading estimates of these bias and error types can arise if *proportional bias* or *heteroscedasticity* are present in the data. However, when a relationship exists between the magnitude of the measurement and difference (*proportional bias* or *heteroscedasticity*) the *fixed bias* and LoA's can be correctly established across the range of measurements using a relatively simple linear regression (OLS) approach. This also allows a quantification of the size of *proportional bias* present. The Bland and

Altman outcome measures are not affected by the range of measurements when a *proportional bias* and *heteroscedasticity* are not present. Like many other agreement measures, they are affected by sample size.

### **3.3.6 Overall agreement statistics conclusions**

There is a rather unusual consensus among biostatisticians that the goal of precision, repeatability and method agreement studies is not to demonstrate agreement, but to detect disagreement or bias. However, biostatisticians disagree, sometimes sharply, on best to achieve this goal. The ultimate goal of the study is to facilitate an objective judgement on the acceptability of a test for daily clinical use. The statistical measurements can evaluate and quantify the type and magnitude error and bias present but cannot provide “yes” or “no” answers on the acceptability of this error. Agreement cannot be statistically significant and should not be accepted based on some absolute, arbitrary, universal statistical criterion such as an ICC CI > 0.75. The magnitude of error (agreement) that is acceptable (tolerable) is ultimately one of clinical judgement (Westgard and Hunt, 1973; Lee et al., 1989; Anastasopoulos, Mergner, Becker and Deecke, 1991; Bland and Altman, 1999; Bruton et al., 2000; Bland and Altman, 2003).

We should ask whether the agreement is good enough for a particular purpose and recognise that agreement (between methods/observers/raters) which may well be good enough for one purpose may not be acceptable enough for another. For example, if a clinician wants to be able to detect small ‘true’ differences between the measurements of interest, the typical variability associated with the measurements must be even smaller. If larger ‘true’ differences are expected, then more variability can be tolerated. The statistical calculations provide the essential data on which this more practical, subjective judgement of acceptability can be made. In order to do this, we need a statistical approach which is sensitive to all error and bias types and at the same time provides a specific and accurate quantification of each. The analysis results (quantification) should also be presented in the units and scale of measurement used in the tests. Ultimately, we ideally want an agreement measurement which is easily understood by users of the measurement method and ultimately easy to interpret with respect to an individual subject or patient (Bland and Altman, 1999).

As highlighted, all approaches have some limitations and can give misleading results when applied inappropriately. Where only two measurements are compared, the Bland and Altman or ordinary least products (OLP) regression analyses are the methods of choice and satisfy all of the above requirements. In addition, both approaches include an extremely important graphical representation of agreement through either a distribution plot or scatterplot. They also provide a method of analysis when the data exhibits a *proportional* or *heteroscedastic* relationship.

It could be argued, that the Bland and Altman logarithmic or regression transformations are significantly more complicated to perform and make the results less easy to interpret and apply. However, these analyses are still easier to execute than OLP regression and the related weighted regression variants (weighted least products - WLP). The more complicated Bland and Altman analyses can also be performed using basic statistical software and spreadsheets; the OLP and WLP analyses do not appear to be widely supported by statistical software.

The additional calculation of an appropriate ICC is not necessary as it provides only a unit free proportional measure adding little to the understanding of the degree of agreement and moreover has considerable statistical limitations and problems in interpretation. However, if more than two measures are compared at once, then the ICC may be the only possibility for getting a measure of at least relative agreement.

So far, it has been assumed that the relationship between one method of measurement and another is linear and the discussion and statistical recommendations have proceeded with this assumption in mind. This is in fact not an unreasonable assumption because it is what usually happens in practice. A non-linear relationship should be picked up by examining a scatterplot or distribution plot and clearly indicates a form of systematic bias.

In subsequent agreement analyses in this thesis, the Bland & Altman approach will be used to assess and quantify the various biases and error types present in the data. The Bland and Altman outcome measures presented will be:



- Distribution plots (difference of two measurements against their mean).
- Mean difference between methods ( $d$  – *fixed bias*) with CI's.
- Limits of agreement (LoA – *random error*) with CI's.
- Modified (regression adjusted)  $d$  & LoA's (*proportional bias & heteroscedasticity*).

## Chapter 4

---

# DEVELOPMENT OF HEAD REPOSITIONING ACCURACY TESTING USING THE ZEBRIS CMS 70P SYSTEM: A STUDY OF METHOD AGREEMENT

---

### 4.1 INTRODUCTION

A few studies have examined the *reliability* characteristics of tests for cervicocephalic kinaesthetic sensibility. These have all focused on the assessment of intra examiner test-retest precision (Table 28 - Chapter 5). As previously highlighted the *reliability* study umbrella has traditionally included studies comparing different methods of measurement (Eliaszew et al., 1994; Ottenbacher, 1995; Bruton et al., 2000; Bland and Altman, 2003; White and Van Den Broek, 2004). The purpose of this study type is to see if the methods agree well enough for one method to replace another or perhaps for the two methods to be used interchangeably (Bland and Altman, 1999; Bland and Altman, 2003). This type of study should be differentiated from concurrent validity. This is where the researcher is interested in whether two instruments measure the same general construct. More specifically, concurrent validity concerns the agreement between the observed value and the true or criterion value of a measure (Ottenbacher, 1995; Hopkins, 2000).

Although studies looking at cervicocephalic kinaesthetic sensibility have used different testing protocols, test repetitions and equipment to measure the parameter of interest, to the author's knowledge there have not been any published studies examining issues of method agreement with these varied approaches. In most studies, the metric used for assessing the kinaesthetic performance is derived from a mean of a number of repetitions (some authors use the term trial instead). Most studies using a SSA (or neutral head position – NHP) repositioning task, particularly the earlier ones, used the mean of 10 repetitions for the measurement error used in the analysis. This seems to have been a rather arbitrary decision. However, some more recent studies have utilised fewer repetitions. The stability (mean value & variance) of these derived measurements may be dependent on the number of trials used to create them (Allison and Fukushima,

2003). This may affect the statistical power for detecting differences between groups and ultimately characteristics of the test (e.g., *reliability*, sensitivity etc.). However, a large number of trial repetitions may be inappropriate for clinical studies and could prove very difficult to complete for individuals with cervical spine pain syndromes. With increasing numbers of repetitions, there is also the extra logistical assessment and analysis burden. There should therefore be a number of repetitions when no more useful information is gained. This would minimise burden on the patients and examiner. It would be useful to investigate the method agreement between the original 10 repetition protocol for the measurement of HRA and a measurement protocol involving fewer repetitions for the mean HRA score.

In establishing a more comprehensive normative knowledge base and evaluating the *reliability* characteristics of HRA-SSA, the current measurement method (laser pointer) also has some limitations. Whilst it utilises a relatively simple equipment design which is inexpensive, easy to execute and may permit a degree of discriminant classification of certain cervicogenic subgroups, the method of measurement and in particular, its subjective and non-remote nature may involve a degree of experimenter bias and inaccuracy. In addition, it does not lend itself to concurrent evaluation of variables such as range of motion and speed of head movement; variables which may have an influence on HRA-SSA. The Zebris CMS 70P system is a relatively new ultrasound based 3-D measuring system that may satisfy many of these methodological requirements and represent a criterion device against which cheaper, low-tech approaches could be compared. The Zebris CMS 70P system has previously been used in cervical range of motion studies (Dvir and Prushansky, 2000; Mannion, Klein, Dvorak and Lanz, 2000; Dvir, Prushansky and Peretz, 2001b; Dvir, Prushansky and Peretz, 2001a) but has only recently been utilised in studies of kinaesthetic sensibility (Lee, Teng, Chai and Wang, 2006; McNair, Portero, Chiquet, Mawston and Lavaste, 2006; Strimpakos, Sakellari, Gioftsos, Kapreli and Oldham, 2006; Demaille-Wlodyka et al., 2007; Lee et al., 2007; Teng et al., 2007). Before commencing the developmental studies for this thesis, no cervicocephalic kinaesthetic studies had been conducted using the Zebris system (Tables 1, 2 & 3 – Chapter 1; Table 28 – Chapter 5).

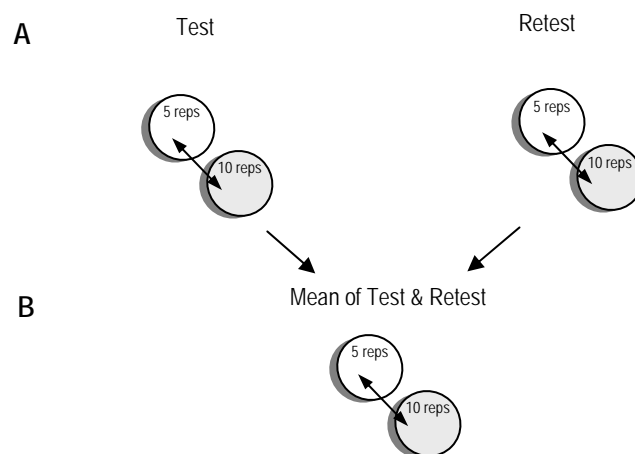
The aim of the present study was to investigate the method agreement in HRA-SSA measurement between a method taking the mean of 5 repetitions of repositioning and one using 10 repetitions

For all of the head repositioning investigations, testing was carried out in the vertical (Flex & Ext  $\Rightarrow$  0 [ $\theta$ X axis]) and horizontal planes (LR & RR  $\Rightarrow$  0 [ $\theta$ Y axis]). However, as the length of a session required to gather the data for the trials in both planes would have been prohibitive (estimated 45 mins), the LR & RR  $\Rightarrow$  0 and Flex & Ext  $\Rightarrow$  0 trials were divided into two separate investigations carried out one month apart. It was also felt that this experimental approach could help minimise the effects of possible confounding variables such as subject fatigue and loss of concentration that have been previously reported. Only the horizontal plane HRA-SSA is reported in this thesis.

## 4.2 METHODS

### 4.2.1 Study Setting and Design

This study took place in the research laboratory at the AECC, Bournemouth, UK. The same room was used on each occasion. The study utilised non-concurrent measurements and was designed to allow the analysis of method agreement with singular comparisons but also with repeated measurements (Figure 28).



**Figure 28.** Diagrammatic representation of method agreement study design and data analysis. A – method agreement at test and retest session (singular); B – method agreement with repeated measurements (mean of test & retest).

## 4.2.2 Subject Recruitment and Selection

The study population was a convenience sample of volunteers drawn from AECC staff, faculty and students. Male and female subjects between the ages of 18 to 55 years were invited to participate after an initial verbal presentation of the nature of the study, the possible risks of participation and the commitment required. Volunteers were then further questioned individually by the investigator regarding their past and current medical history after verbal consent and assurance of confidentiality; if they met any of the exclusion criteria listed in Table 16, they were excluded from participation in the study.

**Table 16.** Inclusion & Exclusion criteria

---

Inclusion
1. Age 18-55.
2. Males and females.
Exclusion
1. Current or previous history cervical spine pain.
2. Prior history of cervical injury or trauma.
3. Current treatment for any musculoskeletal complaint.
4. Persistent or frequent headaches (of any type).
5. Cervical Radiculopathy &/or Myelopathy
6. Inflammatory Arthritis involving C-spine
7. Tumour or infection involving C-spine
8. Vertebrobasilar Artery Insufficiency
9. Neurological disease such as MS, MN, Parkinson's, Syringomyelia etc.
10. History of Dizziness
11. Known congenital anomalies involving the C-spine.
12. Systemic disease such as Diabetes Mellitus.

---

*Abbreviations:*  
C-Spine – Cervical Spine  
MS – Multiple Sclerosis  
MND – Motor Neurone Disease

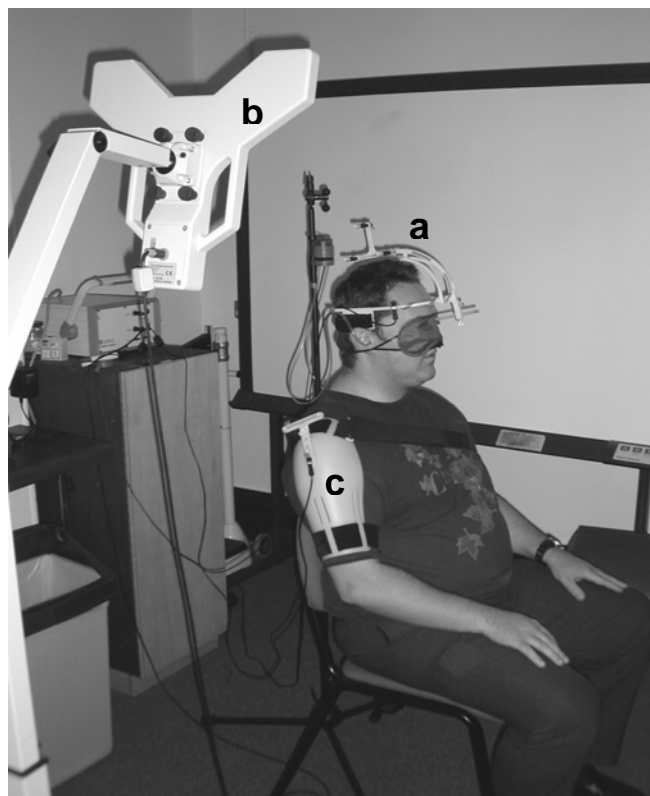
---

All subjects who met the inclusion/exclusion criteria were informed that they would need to avoid manual therapeutic procedures (e.g., massage, stretching, manipulation) applied to the thoracic and cervical spine regions during the two-week period between HRA-SSA measurements. In addition, the chiropractic students were informed that during this period they would need to avoid being subjects for the purpose of teaching

and practising manual procedures (palpation, manipulation set-ups, other examination techniques) involving thoracic and cervical spine regions. Subjects were asked to report if they experienced any neck pain or discomfort during the period between measurements so their data could be omitted from the analysis. They could also discontinue participation at any time, particularly if they experienced significant discomfort before, during or after conducting the trials.

#### 4.2.3 Measurement Instrumentation

Measurement of cervicocephalic kinaesthetic sensibility was performed using the Zebris CMS 70P system (Zebris Medizintechnik GmbH, Isny, Germany). This method of measurement is based on the determination of the spatial coordinates around three orthogonal axes, of miniature ultrasound (US) transmitters. This is accomplished by measuring the sound pulse time delay between the US transmitters and the three microphones housed in the remote measuring sensor. The transmitters are arranged in two triads and attached to head and shoulder plastic frame respectively (Figure 29).

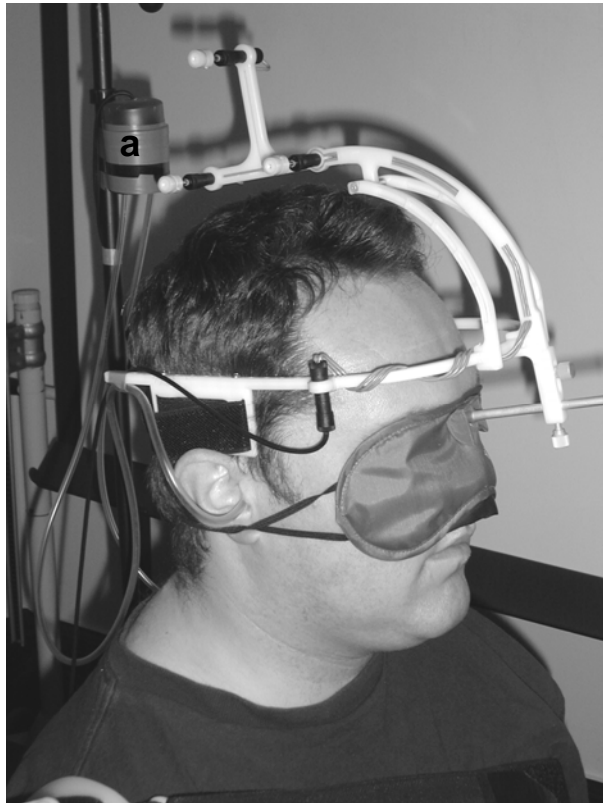


**Figure 29.** Zebris CMS 70P system: a – head attachment, b – remote measuring sensor, c – shoulder attachment.

The shoulder frame acts as a reference transmitter enabling the position of the head with respect to the trunk to be ascertained as opposed to head in space data. The system operates at a maximum sampling rate of 200Hz / number of selected markers. In the present study, the measuring rate was set at 20 Hz / per US marker (six markers; 50ms/sample). The digital 3-D position data was processed by the system's dedicated data acquisition software (WinData) and graphically displayed in real time on a personal computer (PC). It was also saved as a data file that could be exported in a number of formats for offline analysis in other data capture and/or statistical analysis software.

In pilot studies undertaken at the AECC, some subjects reported that they were getting auditory cues for their head position from the 'ticking' noise made by the six US transmitters on the head and shoulder apparatus and the voice of the experimenter when issuing the movement instructions. An earphone system was therefore developed that obliterated the pulsed sound from the transmitters but allowed a bilaterally symmetrical presentation of auditory cues and communication from the experimenter without any additional proprioceptive cues. This system also allowed the experimenter to deliver pre-recorded movement instructions (MPEG files) to the subjects utilising the auditory output facility of the PC used for recording the 3-D data.

Earpieces consisted of a pair of flexible silicone earplugs with the central canal removed to allow sound transmission. One end of a length of 5 mm (external diameter) polypropylene piping was inserted into this central canal of each earplug, attached to the plastic head frame and the other end connected to a small sealed container housing the amplifier (Figure 30). This amplifier was suspended behind the patient using a microphone stand and connected to the auditory output socket of the PC.



**Figure 30.** Earphone system and attachment to the headpiece: a – sealed container housing the speaker.

#### **4.2.4 Measurement protocol**

***Kinaesthetic sensibility testing.*** The investigator conducted all measurement procedures (same examiner as for the first study). All measures were repeated after 14-16 day interval. Measurements took place between 9:00 am and 4:00 pm. As the group of subjects were generally recruited a few weeks prior to the measurement session, on arrival they were checked again with respect to the inclusion/exclusion criteria. They were also reminded of the overall experimental protocol and asked for final verbal consent to participate.



For the HRA-SSA measurements, the subjects were seated in an ordinary fixed seat chair with a backrest for the lumbar and lower thoracic regions only (Figure 29). They were asked to assume a relaxed posture, sit as far back into the chair as possible, let their arms hang down by their sides and place their feet parallel on the ground with the knees at approximately a 90° angle. At this stage, a sleeping mask was then placed over the subjects head but not so that it occluded their vision. The head and shoulder frames were then attached to the subject and checked for comfort. The HRA-SSA testing procedure and objectives was then explained one more time.

Each subject had a similar explanation delivered in a similar manner. The SSA position was described to the subjects as “the subjective position of the head with respect to the trunk that was vertically upright without rotation (anatomical position)”. They were also informed that on hearing the command to move the head, they should perform a maximal rotation of the head, then immediately try to relocate back to the SSA with maximum of precision and then hold this position until they received the next instruction to move. The subjects were also instructed to avoid movement of the trunk during these repositioning tasks. No speed instruction was given.

The Zebris system was then turned on and each subject asked to place the earpieces into the outer ear but only deep enough to occlude the sound of the US transmitters. Clean earpieces were used with each subject. Having completed this, all subjects then had their vision occluded using the sleeping mask for the remainder of the testing procedure. All were instructed to keep their eyes closed behind the mask. The room lights were then switched off.

The test procedure was divided into two parts: 1) determining the SSA and 2) performing the HRA-SSA task. The subject was first asked to find what they perceived as the ‘straight ahead’ (SSA) position of their head with respect to the trunk, as previously defined, and notify the examiner verbally when they felt correctly aligned. The Zebris system was then set to record mode, ‘zeroed’ (calibrated) so that this position was defined as 0° for each subject. The subject was then asked to recognise and remember this position for 2-3 seconds and reminded that this was the SSA or target that they were to try to relocate back to as accurately as possible after each movement in the trial. To initiate the repositioning task (as described earlier), the subject was then simply

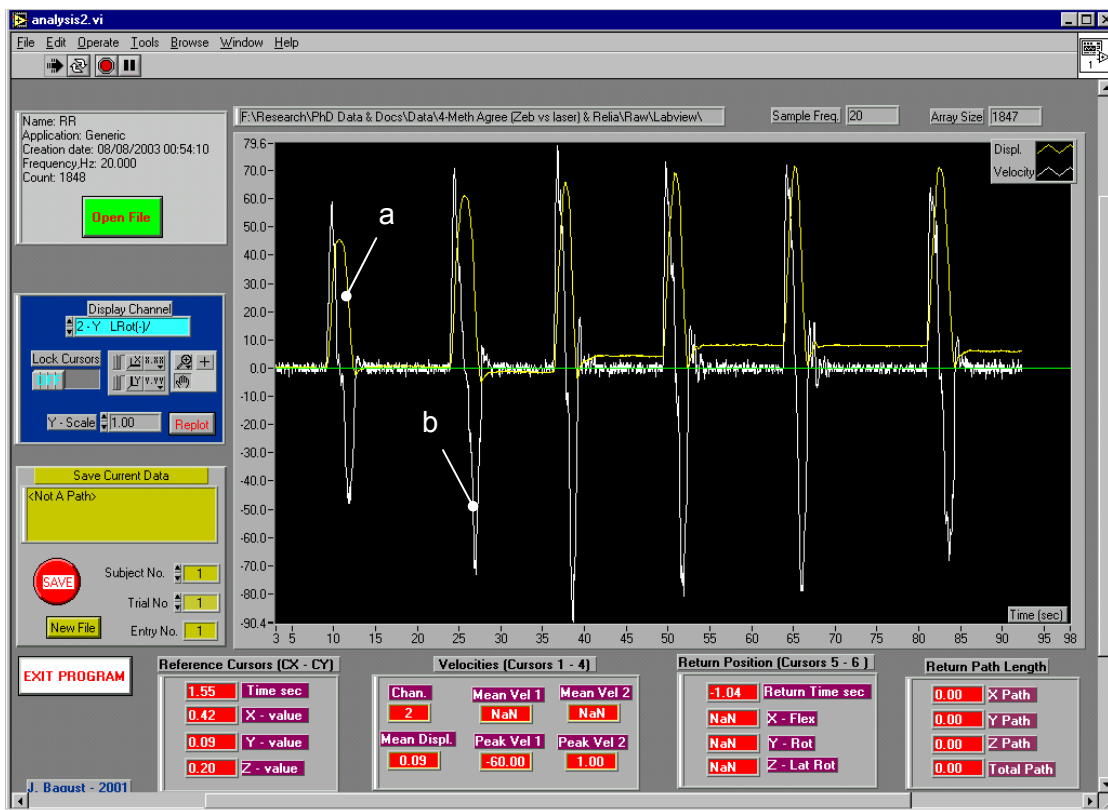
asked to move left (LR  $\Rightarrow$  0) or right (RR  $\Rightarrow$  0) depending on the movement direction under investigation. This instruction was delivered through the earphone system using the MPEG files on the PC. Once the subject had appeared to stop moving at the reposition point, the examiner waited approximately 6-8 seconds before issuing the next instruction to move.

Each movement trial (LR  $\Rightarrow$  0 or RR  $\Rightarrow$  0) consisted of a five and a ten repetition procedure of HRA-SSA. The subjects performed four trials of HRA-SSA (e.g., LR  $\Rightarrow$  0); each of either a five or ten repetition for each movement direction. Approximately two minutes rest was allowed between each trial and the same procedure of subjective alignment to reference zero was carried out at the start of each of these trials. A typical session lasted 20 minutes. The order of the trials within each session (e.g., LR – 5, RR – 10) was varied to avoid any possible effects of order.

At the re-test session, two weeks later, the same protocol was repeated but a different order of trials was used for each subject. The subjects were also checked to see whether they had experienced any neck related complaints or received any manual procedures to the cervico-thoracic region since the last session, and whether there were any other factors that may affect kinaesthetic performance.

#### **4.2.5 Data Processing**

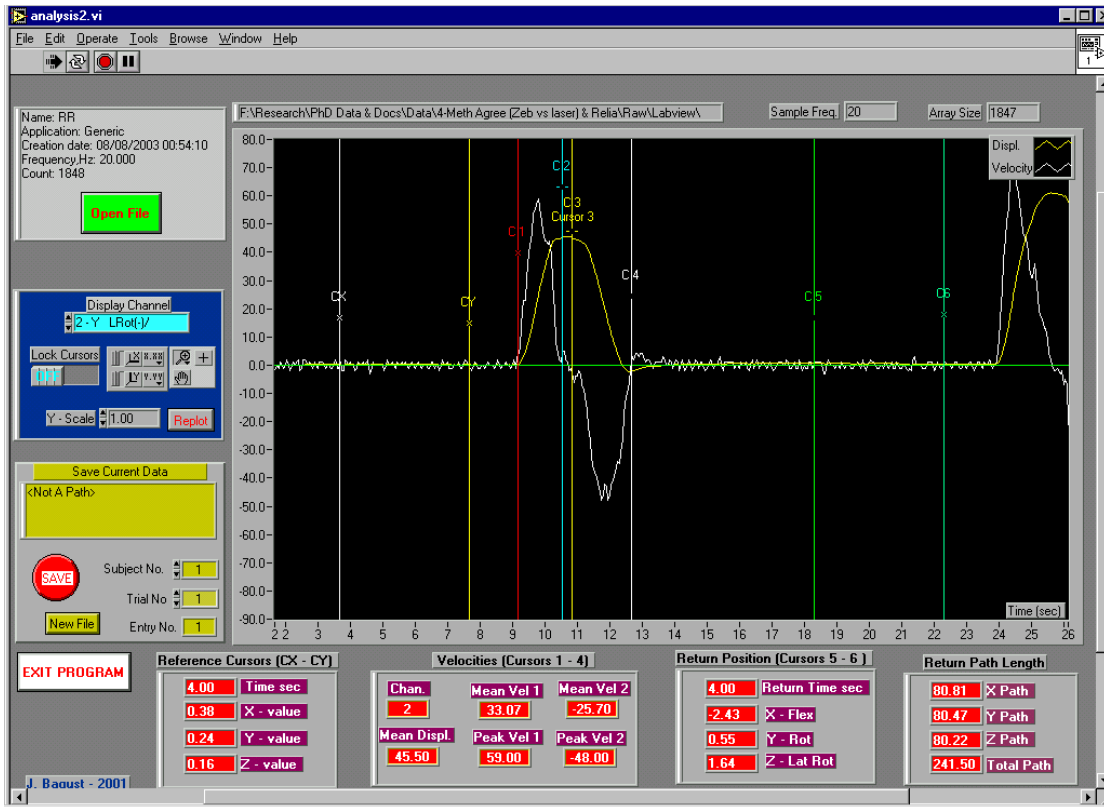
To obtain the values for the kinaesthetic sensibility variables required in these studies, a specific LabVIEW™ (National Instruments Corporation, Austin, TX, USA) application was developed in house by Prof. J. Bagust. The raw angular 3-D position data for each trial recorded by WinData was exported as an ASCII file. The data in this format was then imported into LabVIEW for measurement analysis. The LabVIEW analysis window displays the head on trunk position (degrees) vs time data graphically for the trial selected but only around one of the axes of rotation (Figure 31). However, each axis of rotation can be selected for graphical display and the measurement values for the other two axes of rotation are always displayed at the bottom of the screen in each window selected. The primary axis of rotation associated with the repositioning trial was selected for purposes of measurement analysis (e.g.,  $\theta_Y$  for LR & RR).



**Figure 31.** General view of the LabVIEW™ application measurement analysis window with cursors removed: a – position/time plot representing head movement; b – plot of position/time differential representing head movement velocity

Using the LabVIEW application, the values obtained for each head repositioning repetition were; HRA-SSA, active cervical range of motion (AROM) and means & peak outward & inward angular velocities (Figure 32). AROM & velocity data were not used for the current study but could be useful when making group comparisons as highlighted in Chapter 2. These measurements were obtained by manually placing reference cursors at specific points along the position/time plots of the data (Figure 32). The initial reference zero (CX-CY) and SSA (C5-C6) cursors were placed using a 4 second time separation. Cursors C1 and C4 were placed using the velocity tracing as a guide to clear movement of the head (Figure 32). The measurement values were calculated by formulae that used utilized angular and time data between the cursors or specifically at the point of cursor placement. Measurement values for the other two axes of rotation were simultaneously calculated by the LabVIEW application using the cursors placed on the primary axis as reference points. Once the cursors were placed at the appropriate points along the curve for one repetition, the data was saved. This process was then

repeated for all repositioning repetitions within the particular trial and for all the subjects. For each subject, the data used for descriptive and inferential statistical analysis consisted of the mean value of the repetitions in each trial. These values were calculated using Excel 2000 database software (Microsoft® Corp., Redmond, WA,



USA).

**Figure 32.** First HRA-SSA movement repetition for a RR ⇒ 0 trial. The coloured, labelled (e.g. CX or C2) vertical lines are the cursors used to obtain the HRA, ROM & velocity measurements. Cursors: CX – CY, reference zero for HRA; C1, start of movement away from reference zero; C2-C3, peak ROM; C4, initial cessation of trial movement; C5-C6, SSA measurement.

#### 4.2.6 Data Analysis

The absolute (unsigned) error (AE) and the signed error (SE) were used for comparative analysis of HRA. The AE was derived from the original directional data (SE) using Excel 2000 database software (Microsoft® Corp., Redmond, WA, USA). For each subject, the AE was derived using the ‘ABS’ (absolute) function in Excel. As described in the previous study, the AE is often presented as a logical measure of overall level of HRA performance. This signless error measurement gives information about how close the responses of the performer are to a specific target but without regard to direction.

Zero degree ( $0^\circ$ ) is the best possible score for AE. The SE is a measure of the central tendency of the repositioning performance. As it encompasses the direction of error, it is a reflection of overshoot/undershoot characteristics or directional bias. All data sets (AE and SE) were examined for normality using a combination of statistical testing (Kolmogorov-Smirnov or K-S test) and observation of the data using frequency histograms, Q-Q (normal) plots and box plot summaries of the measures of central tendency (mean and median) and distribution (see results section - Figure 34).

Analysis of method agreement between trials of five or ten repetitions was conducted using the Bland and Altman approach (Altman and Bland, 1983; Bland and Altman, 1986; Bland and Altman, 1999; Bland and Altman, 2003) (Chapter 3 and Appendix III). This approach provides an estimation of the bias and error between two measurement methods and determines the limits within which most of the differences will lie; the 95% limits of agreement (Bland and Altman, 1986). Clinically, this makes it possible to determine whether these two methods of measuring HRA-SSA can be used interchangeably. As stated earlier, the present study was designed to allow the analysis of method agreement (5 vs 10 repetition trials) with singular comparisons but also with repeated measurements (Figure 28). A within-trial subanalysis of 5 vs 10 repetitions method agreement was conducted to investigate the effect of test-retest biases and errors on the method agreement using non-concurrent measurements. For this analysis, the HRA-SSA obtained from the first 5 reps in the 10 rep trials, at the retest session, was compared with the HRA-SSA obtained from the mean of all 10 reps in the trial.

All calculations and distribution plots for the Bland and Altman analysis were produced using Excel 2000 database software (Microsoft<sup>®</sup> Corp., Redmond, WA, USA). All other statistical analyses related to this approach (e.g., statistical and graphical checking of assumptions and regression analysis) were carried out using SPSS for Windows (Rel. 12.0.0. 2003. Chicago: SPSS Inc). The Bland and Altman outcome measures presented are:

- Distribution plots (difference of two measurements against the their mean)
- Mean difference between methods ( $d - \text{fixed bias}$ ) with CI's.
- Limits of agreement (LoA – *random error*) with CI's.
- Modified (regression adjusted)  $d$  & LoA's (*proportional bias & heteroscedasticity*).

## 4.3 RESULTS

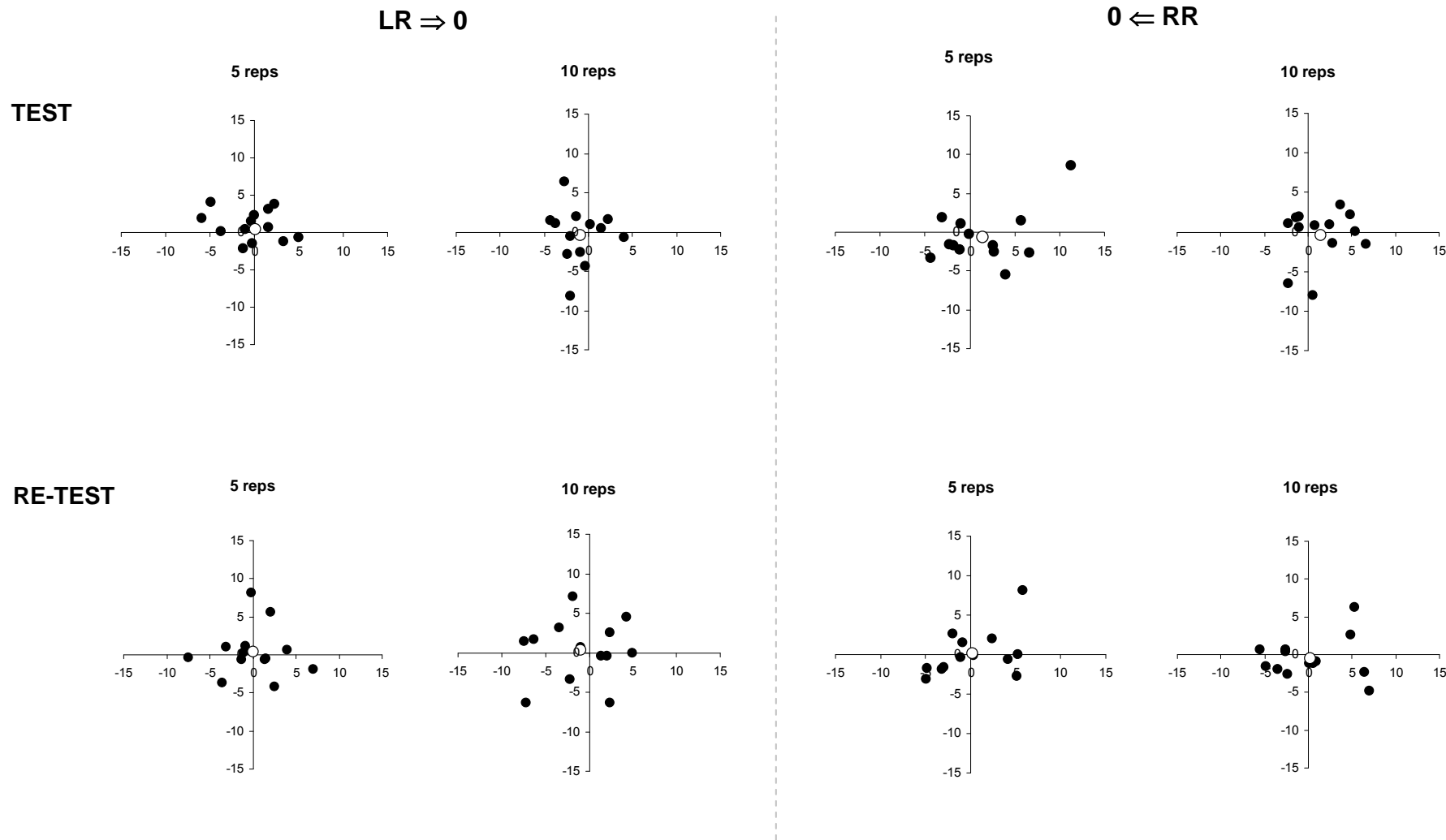
### 4.3.1 Group demographics and characteristics

Twenty healthy subjects (12 men, 8 women) aged 19-52 years (mean age  $\pm$  SD = 28.3  $\pm$  8.3 yrs) agreed and were eligible to participate in the investigation after the initial verbal presentation. However, only thirteen of them performed all the necessary tests over the two-week testing period and provided completed data sets (four subjects could not attend for the initial measurement, one developed neck pain and stiffness before the initial measurement and one did not attend for the second measurement session).

This group consisted of nine men and four women aged 20-48 years (mean age  $\pm$  SD = 27.2  $\pm$  6.3 yrs).

### 4.3.2 Kinaesthetic Sensibility Testing.

An overview of the individual subject HRA-SSA values (mean values of the five and ten repositioning repetitions) for each of the trials at the test and retest sessions, is presented in Figure 33 using scatter plots. Only the  $\theta X$  (vertical) and  $\theta Y$  (horizontal) components are reflected in these plots. A comparison of the distribution and measures of central tendency for the absolute (unsigned) and signed error for the group at each trial is presented in Figures 34 and 35 using box plots. All data sets passed the normality testing using the Kolmogorov-Smirnov test ( $P > 0.1$ ) and were generally considered to exhibit a normal distribution after observation of the graphical summaries. As the sample size was relatively small ( $n < 15$ ), this was based mainly on observation of the Q-Q plots. The descriptive results for the HRA-SSA were therefore summarized using the mean, standard deviation (SD) and 95% confidence intervals for the mean (95% CI) (Tables 17 and 18). The mean absolute HRA-SSA across all trials ranged from 2.1° to 3.6° and 1.8° to 2.9° for the  $\theta Y$  and  $\theta X$  error components respectively. The absolute repositioning errors were generally greatest in the direction of primary motion for the trials ( $\theta Y$  axis). The signed repositioning errors highlighted a tendency for subjects to undershoot the target with both the LR and RR  $\Rightarrow$  0 trials.

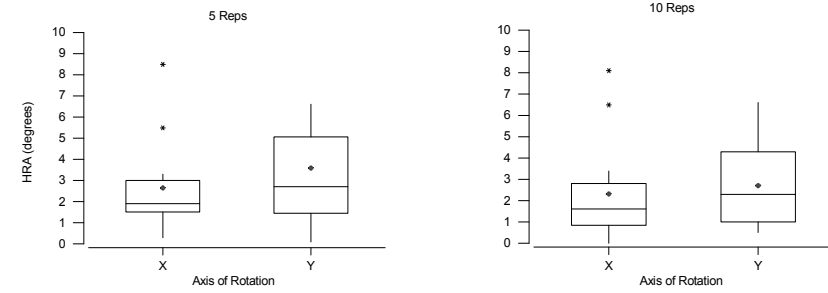
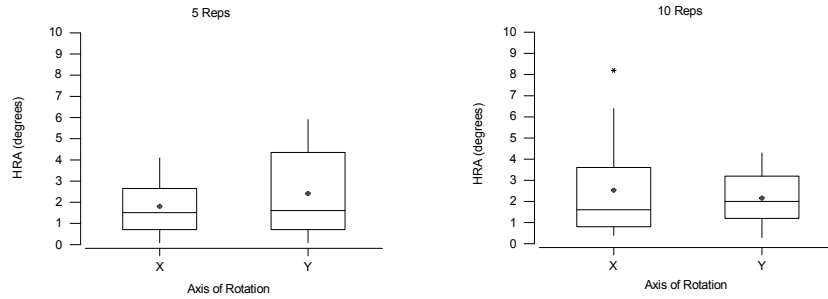


**Figure 33.** Scatter plots showing the individual subject head repositioning accuracy to subjective straight ahead (HRA-SSA ) in the two horizontal rotation directions (L & RR) for each of the trials at the test and retest sessions. All data are presented as degrees. Abbreviations; left rotation (LR  $\Rightarrow$  0), right rotation (RR  $\Rightarrow$  0); the empty circle (O) on the scatter plots is the central point of the data (in plane of motion) represents overshoot/undershoot characteristics in plane of movement).

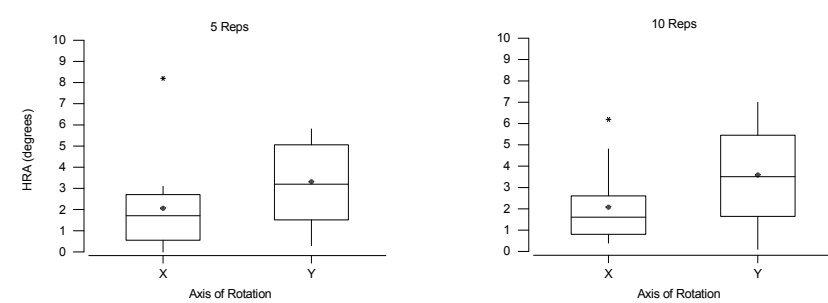
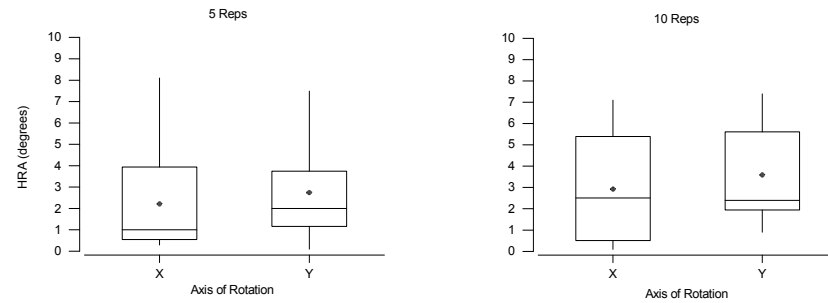
LR  $\Rightarrow$  0

0  $\Leftarrow$  RR

TEST



RE-TEST



**Figure 34.** Box plots of absolute horizontal rotation ( $\theta_Y$ ) and vertical rotation ( $\theta_X$ ) repositioning error (degrees) for each of the trials at the test and retest sessions. All data are presented as degrees. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). For abbreviations, see Figure 33



**Table 17.** Absolute (unsigned) vertical ( $\theta X$ ), horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =13) for trials in the horizontal movement plane (L & RR) conducted with the 5 and 10 repetition testing procedure.

Primary Motion	Axis of rotation	Test				Retest				Combined* (Repeated measures)			
		5 Reps		10 Reps		5 Reps		10 Reps		5 Reps		10 Reps	
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI
LR	$\theta X$	1.8 $\pm$ 1.26	1.04 $\rightarrow$ 2.56	2.5 $\pm$ 2.41	1.09 $\rightarrow$ 4.00	2.2 $\pm$ 2.45	0.72 $\rightarrow$ 3.69	2.9 $\pm$ 2.48	1.44 $\rightarrow$ 4.44	2.0 $\pm$ 1.60	1.04 $\rightarrow$ 2.97	2.7 $\pm$ 2.31	1.35 $\rightarrow$ 4.13
	$\theta Y^\dagger$	2.4 $\pm$ 1.95	1.21 $\rightarrow$ 3.58	2.1 $\pm$ 1.30	1.35 $\rightarrow$ 2.92	2.7 $\pm$ 2.29	1.38 $\rightarrow$ 4.14	3.6 $\pm$ 2.22	2.26 $\rightarrow$ 4.94	2.6 $\pm$ 1.52	1.66 $\rightarrow$ 3.50	2.9 $\pm$ 1.39	2.03 $\rightarrow$ 3.71
RR	$\theta X$	2.6 $\pm$ 2.17	1.34 $\rightarrow$ 3.96	2.3 $\pm$ 2.38	0.89 $\rightarrow$ 3.77	2.1 $\pm$ 2.10	0.79 $\rightarrow$ 3.32	2.1 $\pm$ 1.71	1.05 $\rightarrow$ 3.12	2.4 $\pm$ 1.97	1.16 $\rightarrow$ 3.54	2.2 $\pm$ 1.11	1.54 $\rightarrow$ 2.98
	$\theta Y^\dagger$	3.6 $\pm$ 2.97	1.79 $\rightarrow$ 5.38	2.7 $\pm$ 1.93	1.54 $\rightarrow$ 3.87	3.3 $\pm$ 1.86	2.19 $\rightarrow$ 4.43	3.6 $\pm$ 2.27	2.20 $\rightarrow$ 4.95	3.4 $\pm$ 2.05	2.21 $\rightarrow$ 4.69	3.1 $\pm$ 1.91	1.99 $\rightarrow$ 4.30

Abbreviations: left and right rotation (LR, RR), HRA-SSA – head repositioning accuracy to subjective straight ahead.

\* Mean of test & retest error

$\dagger$  Primary axis of rotation for the repositioning test

No statistically significant differences were demonstrated using a paired *t*-test (two-tailed) between 5 and 10 reps trials at the test and retest sessions and with the combined data.

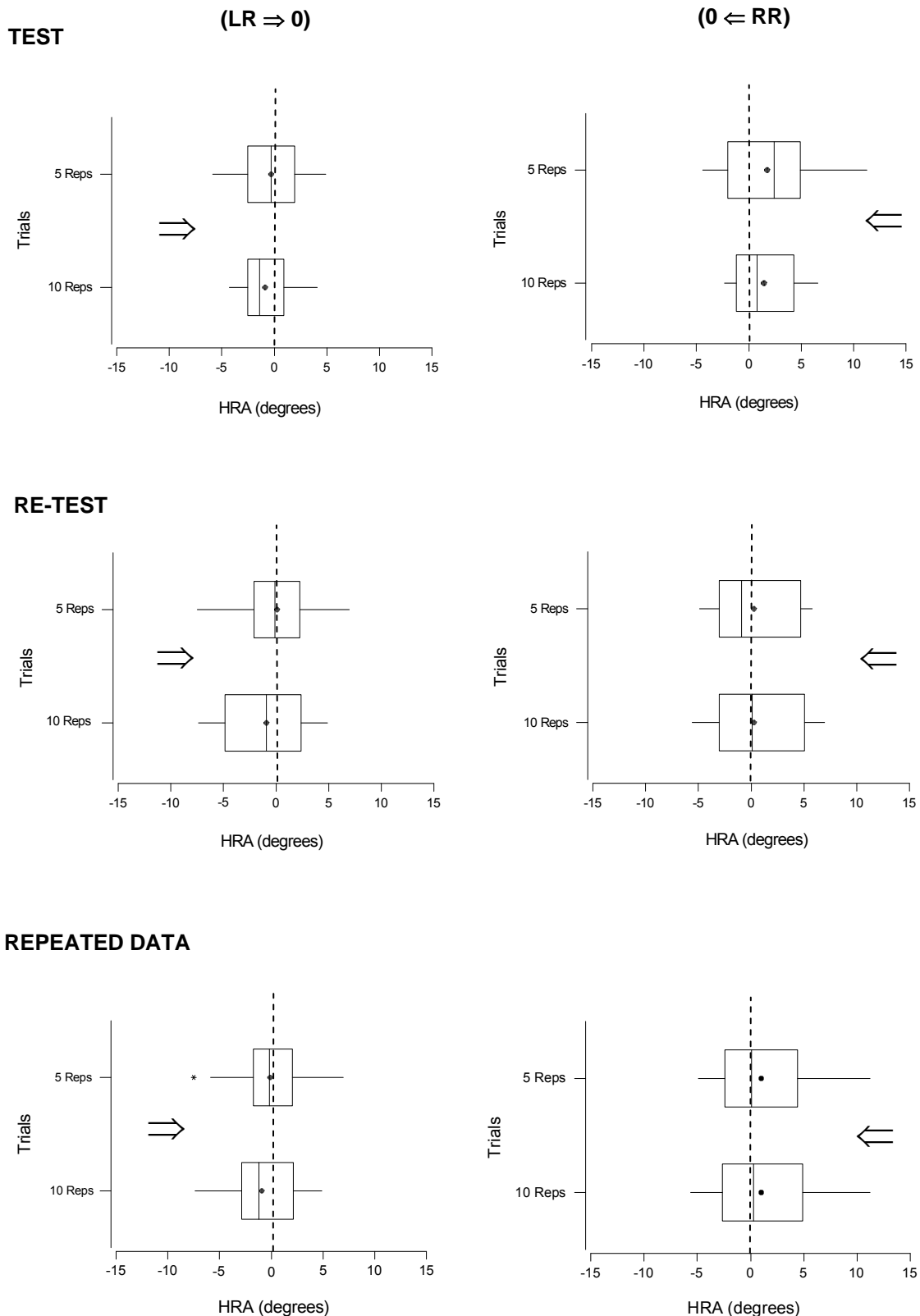
**Table 18.** Signed horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =13) for trials in the horizontal movement plane (L & RR) conducted with the 5 and 10 repetition testing procedure. These data represent the overshoot/undershoot characteristics in the plane of movement for HRA-SSA.

Primary Motion	Axis of rotation	Test				Retest				Combined* (Repeated measures)			
		5 Reps		10 Reps		5 Reps		10 Reps		5 Reps		10 Reps	
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI
LR	$\theta Y$	-0.3 $\pm$ 3.15	-2.19 $\rightarrow$ 2.34	-0.9 $\pm$ 2.40	-2.34 $\rightarrow$ 0.56	0.1 $\pm$ 3.67	-3.47 $\rightarrow$ 1.66	-0.9 $\pm$ 4.24	-2.71 $\rightarrow$ 1.89	-0.1 $\pm$ 3.01	-1.92 $\rightarrow$ 1.72	-0.9 $\pm$ 3.51	-2.80 $\rightarrow$ 0.99
RR	$\theta Y$	1.5 $\pm$ 4.51	-1.23 $\rightarrow$ 4.22	1.5 $\pm$ 3.05	-0.36 $\rightarrow$ 3.32	0.3 $\pm$ 3.90	-2.11 $\rightarrow$ 2.61	0.3 $\pm$ 4.35	-2.37 $\rightarrow$ 2.89	0.9 $\pm$ 3.88	-1.48 $\rightarrow$ 3.22	0.9 $\pm$ 3.25	-1.09 $\rightarrow$ 2.84

Abbreviations: left and right rotation (LR, RR), HRA-SSA – head repositioning accuracy to subjective straight ahead.

\* Mean of test & retest error

No statistically significant differences were demonstrated using a paired *t*-test (two-tailed) between 5 and 10 reps trials at the test, and retest sessions and with the combined data.



**Figure 35.** Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta_Y$ ) in healthy subjects ( $n=13$ ). The middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005).

## Measurement method agreement for HRA-SSA: 5 vs 10 repetitions

For the Bland and Altman analyses, the differences between methods (5 reps – 10 reps) were considered normal in distribution. A visual examination of the distribution plots followed by a statistical evaluation the relationship between the differences and the mean values (Pearson’s  $r$  and absolute residual analysis), indicated the presence of a *proportional bias* and/or *heteroscedasticity* for some of the HRA-SSA error components. For comparative purposes, the Bland and Altman results for all data sets are initially presented using the standard analysis approach (non-regression adjusted).

**Method agreement (5 vs 10 reps): Absolute (unsigned) values.** The results of the standard Bland and Altman analysis of agreement at each testing session and are shown in Table 19 and Table 20. The agreement results using a repeated measurements analysis are shown in Table 21. The distribution plots for the data at the first measurement session indicated a relationship between the differences and the means for the  $\theta X$  error component with both the LR and RR repositioning trials (Figure 37 & 36). For the RR  $\Rightarrow$  0 trial, *heteroscedasticity* was evident but for the LR  $\Rightarrow$  0 trial, both slight *heteroscedasticity* and a more obvious *proportional bias* were present (Figure 36). No relationship was seen with the measurements at the second session. However, for the repeated data analysis *heteroscedasticity* and *proportional bias* were present with the  $\theta X$  error component of RR  $\Rightarrow$  0 trial (Figure 37 & 36). The adjusted bias and agreement estimates (regression analysis) for these data sets are shown in Table 22.

**Table 19.** Results of Bland and Altman analysis for agreement (n=13) between the 5 reps and 10 reps measurement methods at session I for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.74	4.92	0.70	-2.26 → 0.77	-5.67 → 4.18	-8.00 → -3.33	1.85 → 6.52
	$\theta Y^\dagger$	0.26	3.53	0.50	-0.83 → 1.35	-3.26 → 3.79	-4.94 → -1.59	2.12 → 5.46
RR	$\theta X$	0.32	6.90	0.98	-1.80 → 2.44	-6.57 → 7.21	-9.84 → -3.30	3.94 → 10.48
	$\theta Y^\dagger$	0.87	6.06	0.80	-1.00 → 2.74	-5.19 → 6.93	-8.06 → -2.31	4.06 → 9.81

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences

**Table 20.** Results of Bland and Altman analysis for agreement (n=13) between the 5 reps and 10 reps measurement methods at session II for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θX	-0.73	4.43	0.63	-2.10 → 0.63	-5.17 → 3.71	-7.27 → -3.06	1.60 → 5.81
	θY <sup>†</sup>	-0.84	3.48	0.49	-1.92 → 0.24	-4.33 → 2.66	-5.99 → -2.68	1.00 → 4.31
RR	θX	-0.03	3.86	0.63	-1.22 → 1.16	-3.89 → 3.82	-5.71 → -2.06	2.00 → 5.65
	θY <sup>†</sup>	-0.27	4.98	0.70	-1.80 → 1.27	-5.25 → 4.71	-7.61 → -2.88	2.35 → 7.07

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test

**Table 21.** Results of Bland and Altman analysis for agreement using repeated measurements between the 5 reps and 10 reps measurement methods for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>c</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θX	-0.74	4.96	0.70	-2.27 → 0.79	-5.69 → 4.22	-7.71 → -3.68	2.20 → 6.24
	θY <sup>†</sup>	-0.28	4.53	0.64	-1.69 → 1.11	-4.82 → 4.24	-6.44 → -3.20	2.62 → 5.87
RR	θX	0.14	5.04	0.71	-1.41 → 1.70	-4.90 → 5.19	-6.79 → -3.01	3.30 → 7.08
	θY <sup>†</sup>	0.30	5.43	0.77	-1.37 → 1.98	-5.13 → 5.73	-7.26 → -3.00	3.60 → 7.87

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias. s<sub>c</sub> is the adjusted standard deviation of the differences.

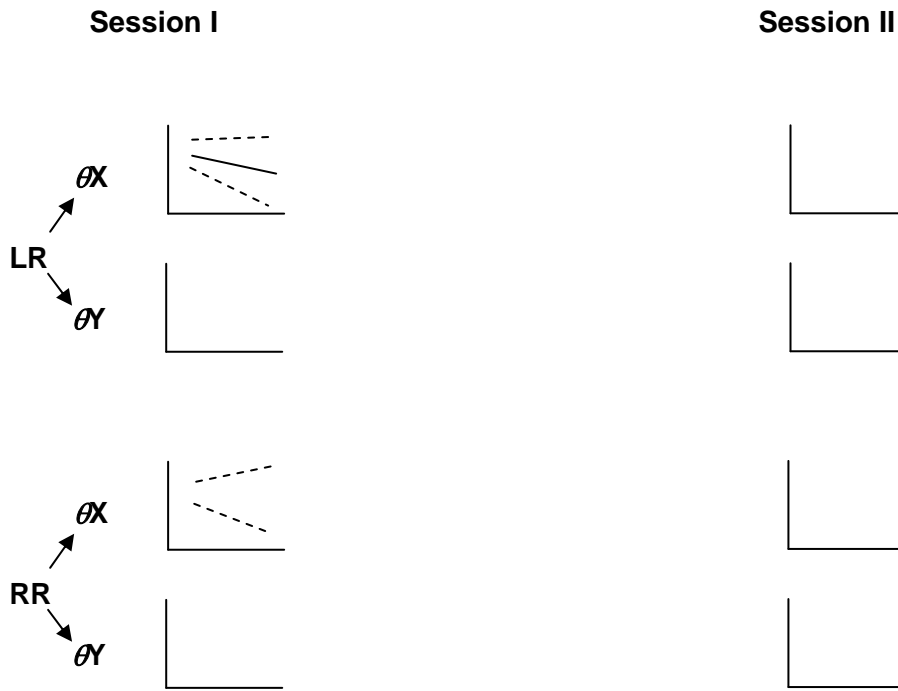
<sup>†</sup> Primary axis of rotation for the repositioning test

**Table 22.** Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=13) between the two different measurement methods (5 reps and 10 reps) with data sets exhibiting a relationship between the mean errors and differences.

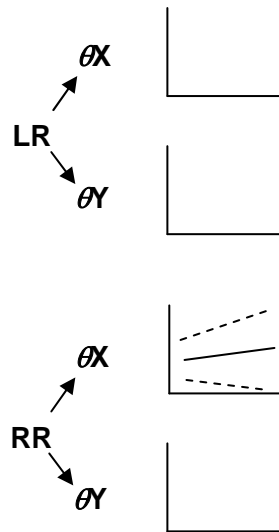
Primary Motion	Session	Axis of Motion	Fixed Bias	95% CI for Fixed Bias	Prop bias	Lowest HRA value	Highest HRA value		
						d ± 1.96s <sub>d</sub>	95% limits of agreement	d ± 1.96s <sub>d</sub>	95% limits of agreement
LR	Test	θX	1.4	-1.01 → 3.84	-1.0	1.3 ± 3.2	-2.9 → 4.5	-3.8 ± 5.0	-8.8 → 1.5
RR	Test	θX	0.3	-1.80 → 2.44	-	0.3 ± 0.6	-0.3 → 0.9	0.3 ± 15.0	-14.7 → 15.3
	Combined	θX	0.8	-3.41 → 0.33	0.7	-0.9 ± 1.6	-2.5 → 0.7	3.0 ± 5.3	-2.3 → 8.3

*Abbreviations:* left and right rotation (LR, RR); Prop - proportional

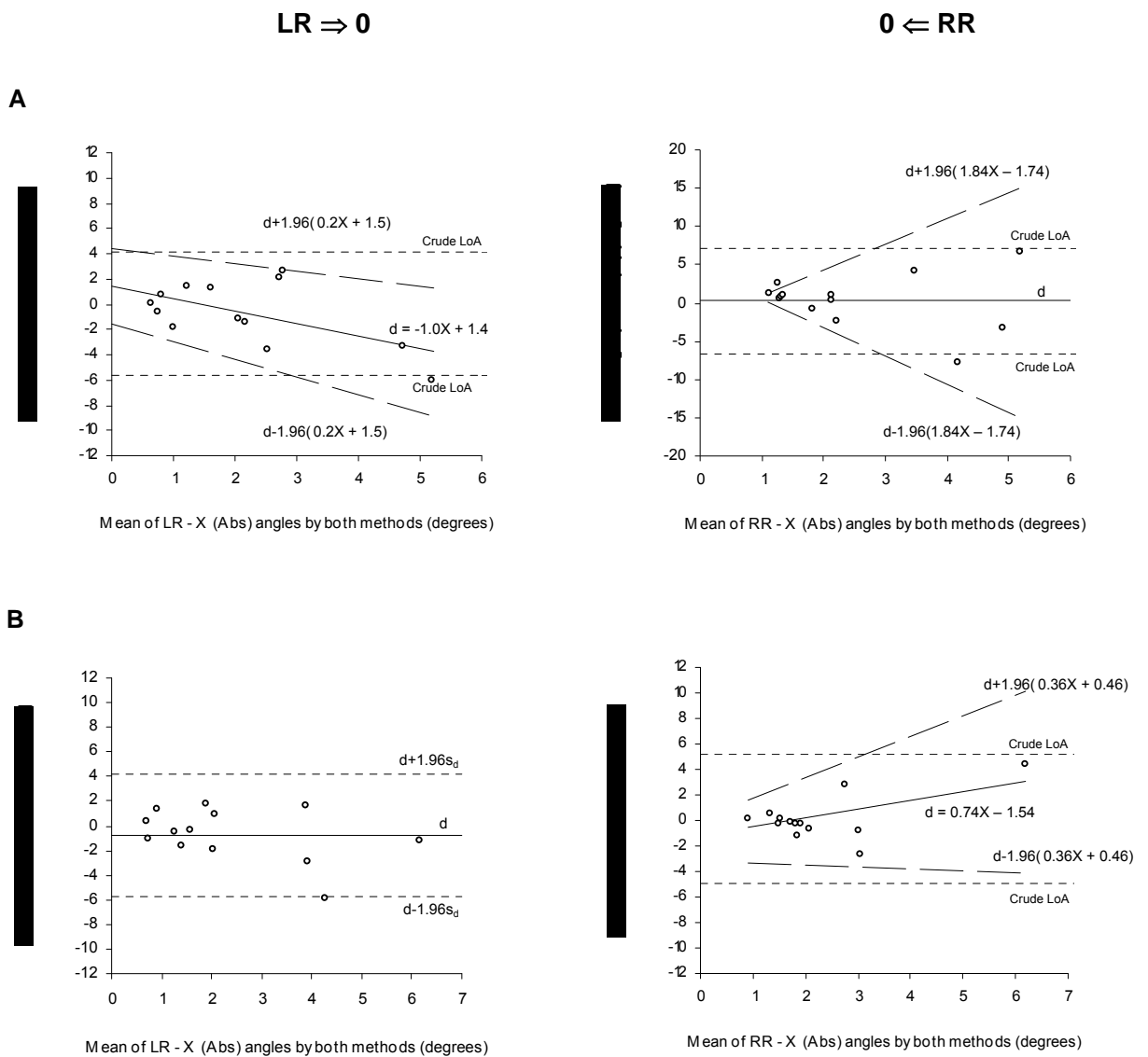
d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps



**Replicate (Session I & II) Measurements**



**Figure 36.** Schematic overview of LR & RR data sets for the Bland & Altman method agreement (5 vs 10 reps trials) that exhibit *Heteroscedasticity* &/or *Proportional Bias*. These errors & biases are only seen for the  $\theta X$  components (perpendicular to plane of movement). The blank graphs represent method agreement data without *Heteroscedasticity* &/or *Proportional Bias* (ie., no relationship between differences and means)



**Figure 37.** Plots of the difference against mean for the absolute 5 and 10 repetitions X-axis head repositioning accuracy (HRA) at the first testing session (A) and with replicate measurements (B). On each plot, *fixed bias* is indicated by the mean value line (d); 'limits of agreement' (LoA) are given by the  $d \pm 1.96sd$  lines.

Overall, the standard and regression adjusted results indicate that there was no consistent and statistically significant *fixed bias* ( $d$ ) between the two methods of measurement although for the  $LR \Rightarrow 0$  data and all data at the retest session, the errors were slightly larger from the mean of the 10 repetition data. With the majority of measurement comparisons (11 of 12), the mean difference ( $d$ ) between measurement methods was less than  $1^\circ$ . Estimates of *random error* ( $\pm 1.96s_d$ ) using the standard Bland and Altman analysis ranged from 4.4 to 6.9 degrees for the  $\theta X$  errors and 3.5 to 6.1 degrees for the  $\theta Y$  errors. The  $\theta Y$  *random error* for the  $RR \Rightarrow 0$  trial was consistently higher than those seen with the  $LR \Rightarrow 0$  trials. The incorporation of test-retest variance did not in general result in consistently larger estimates of *random error*. However, when *proportional bias* and/or *heteroscedasticity* was evident, the regression adjusted  $\theta X$  *random error* estimates ranged from 0.6 to 3.2 degrees for lowest mean HRA measurements and 5.0 to 15.0 degrees with the largest mean measurements of  $5^\circ$ . For the two  $\theta X$  data sets exhibiting a *proportional bias*, the direction was inconsistent (Table 22). Using the 10 repetition method data as a reference, this represents proportionally larger 10 rep values with the  $LR \Rightarrow 0$  trial and proportionally smaller values for the  $RR \Rightarrow 0$  trials with repeated measures. The overall impact of the *proportional bias* on the adjusted Bland and Altman results was a positive or negative shift in the limits of agreement (LoA's) with increasing size of measurements. As previously mentioned, the *fixed bias* estimates were not statistically affected by the *proportional bias*. The presence of *heteroscedasticity* resulted in narrower LoA's for low HRA measurements but wider LoA's for the highest values.

The results of the agreement analysis for the mean of the first five repetitions and all ten repetitions within each trial are shown in Table 23. This represents data from the second measurement session. For all data sets, there was no relationship between the differences and the mean values. An examination for *fixed bias* ( $d$ ) showed that there was a small but consistent difference between the methods with slightly larger errors from the 10 repetition data. This was more evident with the  $LR \Rightarrow 0$  trial and represents a slight increase in repositioning error with increasing repetitions (Figure 38). The mean differences were less than  $0.5^\circ$  for all error components and were not statistically significant. Estimates of *random error* ( $\pm 1.96s_d$ ) ranged from 1.3 to 1.5 degrees for the  $\theta X$  errors and 1.4 to 1.7 degrees. There was a tendency for slightly larger  $LR \Rightarrow 0$  *random error* estimates but the differences were less than 0.3 degrees.

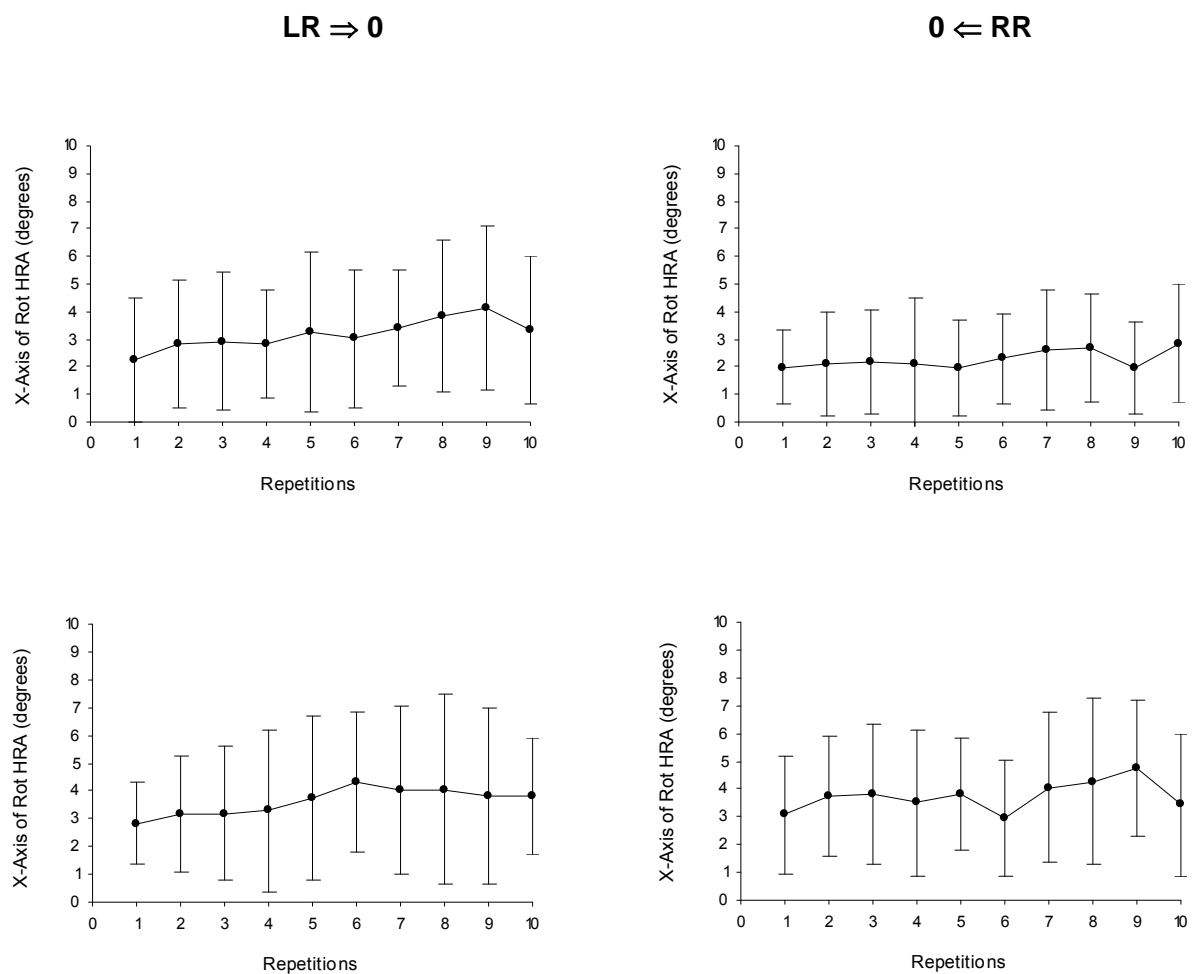
**Table 23.** Results of Bland and Altman analysis for method agreement between two different measurement methods (5 reps and 10 reps within trial) for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.36	1.45	0.21	-0.81 → 0.09	-1.82 → 1.10	-2.51 → -1.13	0.41 → 1.79
	$\theta Y^\dagger$	-0.46	1.71	0.24	-0.99 → 0.06	-2.16 → 1.24	-2.97 → -1.36	0.43 → 2.05
RR	$\theta X$	-0.11	1.27	0.18	-0.50 → 0.29	-1.39 → 1.17	-1.99 → -0.78	0.57 → 1.78
	$\theta Y^\dagger$	-0.11	1.35	0.19	-0.53 → 0.31	-1.46 → 1.25	-2.11 → -0.82	0.60 → 1.89

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences

† Primary axis of rotation for the repositioning test



**Figure 38.** Left & right rotation (LR, RR) absolute (unsigned) head repositioning accuracy (HRA) for each of the 10 repetitions at the retest session. Data are represented as Mean ± SD (degrees).



**Method agreement (5 vs 10 reps): Signed values.** The results of the standard Bland and Altman analysis of agreement at each testing session and are shown in Table 24 & 25. The agreement results using a repeated measurements analysis are shown in Table 26. For all data sets, there was no relationship between the differences and the corresponding mean values. An examination for *fixed bias* ( $d$ ) showed that there was a relatively small but consistent difference between the methods for LR  $\Rightarrow$  0 trials with slightly larger estimates of undershoot from the ten repetition trials. The mean differences were less than  $1.0^\circ$  for all error components and were not statistically significant. The *random error* ( $\pm 1.96s_d$ ) estimates were generally larger for the RR  $\Rightarrow$  0 trials with values ranging from 4.2 degrees to 5.1 degrees for LR  $\Rightarrow$  0 and 5.0 degrees to 8.5 degrees for RR  $\Rightarrow$  0. The incorporation of test-retest variance resulted in *random error* estimates of 5.1 degrees for LR  $\Rightarrow$  0 and 6.9 degrees for RR  $\Rightarrow$  0 (Table 26).

**Table 24.** Results of Bland and Altman analysis for method agreement using between two different measurement methods (5 reps and 10 reps) at session I (test) for the signed Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta Y$	0.60	4.17	0.59	-0.68 $\rightarrow$ 1.89	-3.57 $\rightarrow$ 4.78	-5.55 $\rightarrow$ -1.59	2.80 $\rightarrow$ 6.76
RR	$\theta Y$	0.01	8.25	1.17	-2.54 $\rightarrow$ 2.56	-8.25 $\rightarrow$ 8.27	-12.17 $\rightarrow$ -4.33	4.35 $\rightarrow$ 12.19

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences

**Table 25.** Results of Bland and Altman analysis for method agreement using between two different measurement methods (5 reps and 10 reps) at session II (retest) for the signed Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta Y$	0.99	4.39	0.62	-0.36 $\rightarrow$ 2.34	-3.40 $\rightarrow$ 5.38	-5.48 $\rightarrow$ -1.31	3.30 $\rightarrow$ 7.47
RR	$\theta Y$	-0.01	5.02	0.71	-1.56 $\rightarrow$ 1.54	-5.03 $\rightarrow$ 5.01	-7.41 $\rightarrow$ -2.65	2.63 $\rightarrow$ 7.39

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences

**Table 26.** Results of Bland and Altman analysis for method agreement using repeated measurements between two different measurement methods (5 reps and 10 reps) for the signed Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96S <sub>c</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta Y$	0.80	5.06	1.49	-0.76 → 2.36	-4.26 → 5.86	-6.07 → -2.45	4.05 → 7.67
RR	$\theta Y$	-0.01	6.94	0.98	-2.14 → 2.14	-6.94 → 6.94	-9.49 → -4.40	4.40 → 9.48

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias. s<sub>c</sub> is the adjusted standard deviation of the differences.

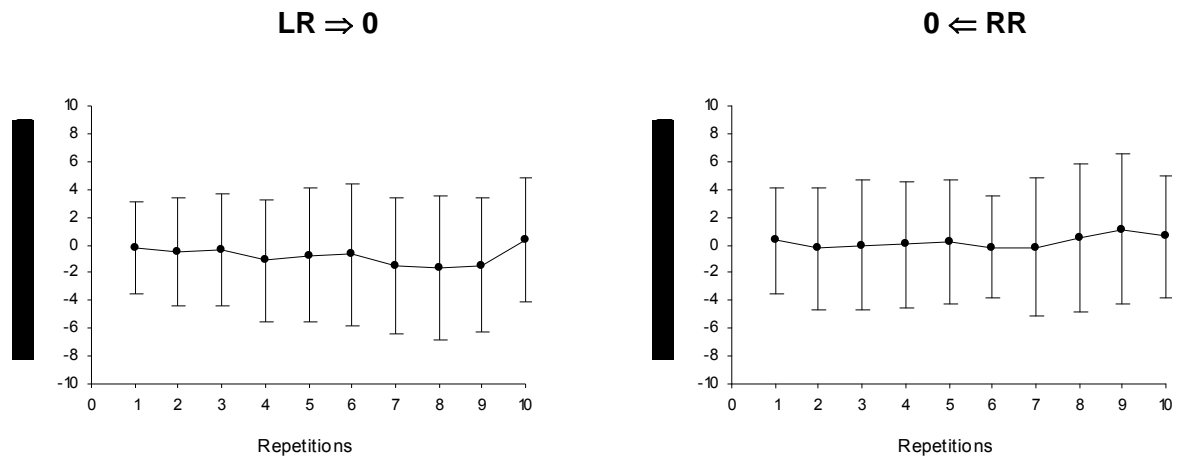
The results of the Bland and Altman analysis of agreement between the signed HRA-SSA from the mean of the first five repetitions and the mean of all ten repetitions within each trial are shown in Table 27. There was no relationship between the differences and the corresponding mean values for both data sets. There was also no consistent and statistically significant *fixed bias* (d) between the two methods and no evidence of a change in repositioning error with increasing repetitions (Figure 39). The mean error calculation for differences (*fixed bias*) was less than 0.5° for both trials (Table 27). Estimates of *random error* ( $\pm 1.96s_d$ ) were markedly less than those values seen with the between trial analysis (1.4° to 1.8° vs 4.2° to 8.5°).

**Table 27.** Results of Bland and Altman analysis for method agreement between two different measurement methods (5 reps and 10 reps within trial) for the signed Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Motion	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta Y$	0.32	1.82	0.26	-0.24 → 0.89	-1.51 → 2.15	-2.38 → -0.64	1.28 → 3.02
RR	$\theta Y$	-0.22	1.41	0.20	-0.65 → 0.22	-1.63 → 1.20	-2.31 → -0.96	0.53 → 1.87

*Abbreviations:* left and right rotation (LR, RR); LoA – Limits of Agreement

d is the mean of differences (5 reps - 10 reps means) ; a +ve value indicates 5 reps > 10 reps; d represents fixed bias; s<sub>d</sub> is the SD of the differences



**Figure 39.** Left & right rotation (LR, RR) signed head repositioning accuracy (HRA) for each of the 10 repetitions at the retest session. Data are represented as Mean  $\pm$  SD (degrees).

### 4.3.3 Summary of Kinaesthetic Sensibility Primary Results.

#### Method agreement for 5 vs 10 reps HRA-SSA.

- **Absolute HRA - Between trials:**
  - 3 of the 12 data sets exhibited *heteroscedasticity* &/or *proportional bias*. These were all for  $\theta X$  error components (perpendicular to plane of movement).
  - No consistent and statistically significant *fixed bias* (d) between the two methods of measurement using the standard and regression adjusted results.
  - Estimates of *random error* ( $\pm 1.96s_d$ ) ranged from 4.4 to 6.9 degrees for the  $\theta X$  errors and 3.5 to 6.1 degrees for the  $\theta Y$  errors.
  - When *proportional bias* and/or *heteroscedasticity* was evident, the regression adjusted  $\theta X$  *random error* estimates ranged from 0.6 to 3.2 degrees for lowest mean HRA measurements and 5.0 to 15.0 degrees with the largest mean measurements of  $5^\circ$ .
  - The incorporation of test-retest variance did not result in consistently larger estimates of *random error* or a change in estimates of *fixed bias*.

- ***Absolute HRA - Within 10 Rep trial (Session II):***
  - No statistically significant *fixed bias* ( $d$ ) between the two methods of measurement.
  - Estimates of *random error* ( $\pm 1.96s_d$ ) ranged from 1.3 to 1.5 degrees for the  $\theta X$  errors and 1.4 to 1.7 degrees for the  $\theta Y$  errors.
  
- ***Signed HRA (overshoot/undershoot) - Between trials:***
  - No statistically significant *fixed bias* ( $d$ ) between the two methods of measurement.
  - Estimates of *random error* ( $\pm 1.96s_d$ ) ranged from 4.2 degrees to 5.1 degrees for LR  $\Rightarrow$  0 and 5.0 degrees to 8.5 degrees for RR  $\Rightarrow$  0.
  - The incorporation of test-retest variance resulted in *random error* estimates of 5.1 degrees for LR  $\Rightarrow$  0 and 6.9 degrees for RR  $\Rightarrow$  0.
  
- ***Signed HRA (overshoot/undershoot) - Within 10 Rep trial (Session II):***
  - No consistent and statistically significant *fixed bias* ( $d$ ) between the two methods of measurement.
  - Estimates of *random error* ( $\pm 1.96s_d$ ) ranged from 1.4 degrees for LR  $\Rightarrow$  0 to 1.8 degrees for RR  $\Rightarrow$  0.

## 4.4 DISCUSSION

The purpose of this study was to see if the methods of measurements using 5 and 10 repetitions for a mean HRA score agree well enough for one method to replace another or perhaps for the two methods to be used interchangeably. As discussed in Chapter 3, it is not possible to say how small limits of agreement should be to represent ‘fair’, ‘good’ or ‘excellent’ agreement. The questions we have to ask is what error types are present and whether the measures agree sufficiently well, i.e., whether the largest likely difference is small enough for the particular purpose for which we want the measurements. This may be different for different purposes (Bland and Altman, 1986; Bland and Altman, 2003).

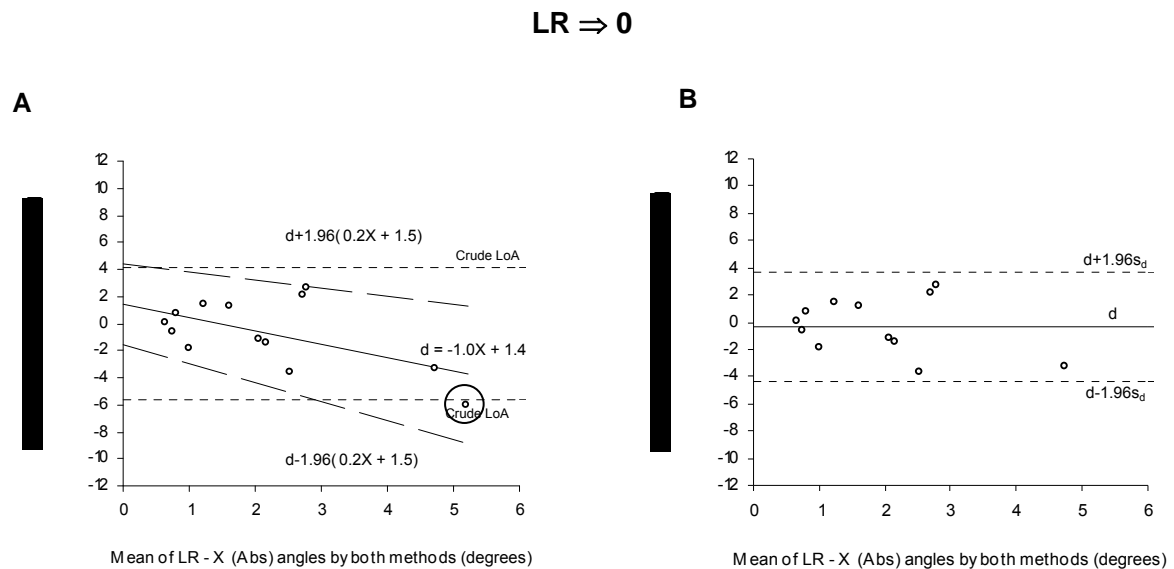
The overall results from this study suggest that the two methods of measurement agree sufficiently well for the 5 repetition method to replace the method using 10 repetitions of repositioning to obtain a mean HRA score and that both could be used interchangeably. In discussing this further, a closer examination of the error and bias types and is needed.

### **Absolute head repositioning accuracy (HRA-SSA)**

If the absolute data is considered, the results reveal that there was no significant or consistent *fixed bias* between the two methods of measurement. This means that when comparing the 5 and 10 repetition methods, there was no constant or additive difference ( $Y = X + a$ ) between the mean scores obtained with each method.

For 3 of the 12 comparisons made (Figures 36 and 37), *heteroscedasticity* of the data were exhibited, two of these also displaying a *proportional bias* (multiply bias;  $Y = bX$ ) as well. These all appeared in the  $\theta X$  axis of rotation, perpendicular to the primary plane of trial movement. This could represent a relative lack of concentration by the subjects on this plane of positioning. However, there was no clear pattern or consistency to the appearance of these errors ie they only appeared at the first testing session and only in the  $RR \Rightarrow 0$  data set when test-retest variance was added in. This rather inconsistent appearance of error and bias may be related to the relatively small sample size and therefore relatively large effect of a small number of outlying scores. If the boxplots of the trials are considered (Figure 34), there are outlying values for only the  $\theta X$  components of repositioning. This is particularly

apparent for the  $RR \Rightarrow 0$  data set at session I. The effect of one outlier on the Bland and Altman analysis is shown in Figure 40. When one outlying data point is removed, the *heteroscedastic* and *proportional bias* relationship disappears and the limits of agreement become narrower and more symmetrical.



**Figure 40.** Bland and Altman Distribution plots for the  $LR \Rightarrow 0$   $\theta X$ -axis method agreement at session I. (A) all data. (B) effect with one outlier removed (outlier is circled in plot A)

If we assume that there is no true *heteroscedasticity* or *proportional bias* in the data and the appearance of these errors/bias in the  $\theta X$  axis data is due to the impact of outlying data on a small sample size, then the only true variance between the methods of measurement is *random* in nature. If we consider the  $\theta Y$  components of the  $LR \Rightarrow 0$  trial, the results suggest that for an individual subject tested with each of the 5 and 10 repetition methods, we can expect the scores using the two methods to vary randomly by up to  $\pm 3.5$  to  $4.5$  degrees (Tables 19 to 21; *random error* =  $1.96S_d$  or  $1.96S_c$ ). For  $RR \Rightarrow 0$  we can expect the scores to vary randomly by  $5.0$  to  $6.1$  degrees (Tables 19 to 21; *random error* =  $1.96S_d$  or  $1.96S_c$ ). It is possible that there is truly a larger *random error* when testing right rotation but this may again be related to the effect of the outlying data within a relatively small sample size.

As introduced in Chapter 3, the sources of variability (*random error*) that may have been captured in this study include the following components.

- Instrument error/variability (*i.e. the variability of the measurement device*).
- Rater error/variability (*i.e. the variability of the researcher/s or clinician/s administering the measurement device*).
- Response variability (*i.e. the variability/stability of the variable/construct being measured*).

In most studies the isolation of rater from instrument and response variability is often not possible. The most robust method agreement study design involves a repeated measures design as this will always be present when methods are interchanged or compared (Bland and Altman, 1999; Berthelsen and Nilsson, 2006). In this study there were essentially two test-retest components; 1) immediate (testing using the 5 and 10 rep methods immediately after each other and 2) two week repeated measures (same measurement protocol at session II as session I). If the  $\theta Y$  axis errors are considered again, the introduction of between week test-retest variance did not result in larger estimates of *random error* compared with the method agreement analysed at each session (Tables 19 to 21). This would suggest that the immediate and between week variance are truly random and approximately equal in size. The immediate test-retest effect can be quantified if the between trial (separate 5 vs 10 rep trials) and within trial (mean of first 5 reps of 10 rep trial vs mean of 10 reps) are compared. If the within trial  $\theta Y$  data are considered, the *random error* for LR  $\Rightarrow$  0 is reduced to  $\pm 1.3$  to 1.5 degrees compared with  $\pm 3.5$  to 4.5 degrees with the between trial analysis suggesting a test-retest *random error* of approximately  $\pm 1.8$  to 2.8 degrees. For the RR  $\Rightarrow$  0 trials, the test-retest *random error* can be estimated at  $\pm 3.5$  to 4.6 degrees. The sources of variability for the within trial method agreement cannot be isolated beyond those highlighted above, namely; instrument, rater and biological variability.

### **Signed head repositioning accuracy (HRA-SSA) – over/undershoot characteristics**

For the signed data, there was no significant *fixed* or *proportional bias* or *heteroscedasticity*. The only variance between the methods of measurement was *random error*. As with the absolute data, the introduction of between week test-retest variance did not result in larger estimates of *random error* compared with the method agreement analysed at each session (Tables 24 to 26). This would again suggest that the immediate and between week variance are approximately equal in size. In essence then, the results show that if an individual subject was tested with each of the 5 and 10 repetition methods, we

can expect the over/undershoot estimates to vary randomly by up to  $\pm 5.1$  to  $6.9$  degrees (for L & RR using the combined data). With the relatively small over/undershoot score seen in this study (Table 18) this means that using the two methods a subjects over/undershoot score may randomly vary from a relative overshoot to undershoot position or vice versa. This may well predominantly represent biological variability between the testing sessions. If the within trial signed data is considered (no between method test-retest variance) the random error is reduced to  $\pm 1.4$  degrees for LR  $\Rightarrow 0$  and to  $\pm 1.8$  degrees for RR  $\Rightarrow 0$ . This suggests a test-retest random error of  $\pm 3.7$  to  $5.1$  degrees. Removing this variable also means that it is less likely that the subjects estimate of over/undershoot characteristics will change when the 5 repetitions method is used for the testing trials.

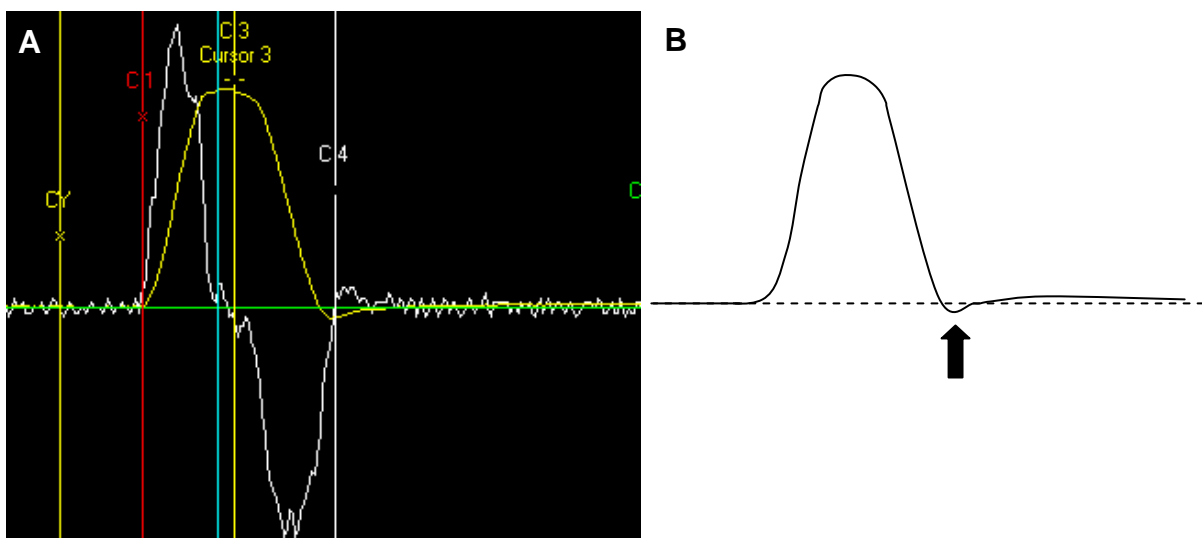
### **Comparison of results with other studies.**

Only one other study has looked at the effect of repetitions of head repositioning on the HRA testing protocol (Swait, Rushton, Miall and Newell, 2007). More specifically, these investigators studied the ‘stability’ and within and between day *reliability* for a HRA-SSA testing protocol. For most subjects, ‘stable’ estimates were derived from 6 or more repetitions (differences between RMSE, SD of the mean and CV of mean RSE values being zero). For *reliability* estimates, the ICC (2,k) was used. The results showed that with a lower bound of the CI for the ICC set at 0.4, the ICC values obtained exceeded this with 5 or more repeats. Although it is not clear whether a *consistency & absolute* variation of ICC was used, the choice of ICC model (2,k) was appropriate. However, as previously discussed (Chapter 3), the use of the ICC value in isolation does not provide any indication of the magnitude of disagreement between measurements. Without this, it is not possible to judge whether the agreement (*reliability*) is good enough for any particular purpose. Although early HRA studies used 10 repetitions to arrive at a mean score for each subject (Revel et al., 1991; Revel et al., 1994; Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998; Heikkila et al., 2000; Humphreys and Irgens, 2002), many subsequent studies have used only 3 repetitions (Chapter 1: Tables 1-3). The results from Swait et al., (2007) suggest that a mean score from 3 repeat measurements may not give the most ‘stable’ and precise results. Further studies are needed to evaluate the testing characteristics of 3 repetition protocol against those using 5, 6 or 10 repeat measurements.



## Observations of HRA characteristics

In the previous investigation (Chapter 2), a short-lived wobble or fine drift was seen in many subjects immediately on reaching the repositioning target followed by a stable control of the head position. This movement has also been noted by other researchers (Humphreys and Irgens, 2002; Feipel et al., 2006; Palmgren et al., 2006). Using the Zebris measuring device, this phenomenon can be clearly seen (Figure 41). On return from the maximal rotation of the head, there is a slight overshoot followed by a correction and then a steady phase of head positioning until the next instruction to move. This indicates that a short latency (approx 2 second) is appropriate before measuring the HRA so as to allow a more accurate assessment of the subject's true position of relocation.



**Figure 41.** A single HRA repetition position time plot for a single subject. (A) the Labview™ application measurement analysis window with cursors; position-time plot represented by yellow trace. (B) a plot of the same position-time plot. The dashed horizontal line represents the reference zero (straight ahead position) calibrated when the subject finds their initial SSA position. The arrow (in B) represents the initial overshoot when repositioning back to SSA.

## 4.5 CONCLUSION

The overall results from this study suggest that the two methods of measurement agree sufficiently well for the 5 repetition method to replace the method using 10 repetitions of repositioning to obtain a mean HRA score and that both could be used interchangeably. Further studies of method agreement for 3 and 5 repetition trials are indicated. It would also be interesting to investigate the optimum sample size needed to obtain a stable estimate of *random error* using the Bland and Altman analysis.

## Chapter 5

---

# INTER & INTRA EXAMINER 'RELIABILITY' AND METHOD AGREEMENT OF THE ZEBRIS AND LASER POINTER METHODS FOR MEASURING HEAD REPOSITIONING ACCURACY

---

### 5.1 INTRODUCTION

As highlighted in Chapters 3 and 4, establishing the characteristics is a vital step in the development and adoption of a new clinical test (Table 8). These characteristics help determine its relative usefulness and which of the functions the test is best suited to. For the HRA measurement parameter and measuring methods, there is currently a specific fundamental need for adequate evaluation of the clinical characteristics. In particular there is a need to establish the *reliability* of the testing procedures and the level of agreement the different methods of measurement (method agreement studies). An important first step in establishing the usefulness and efficacy of any assessment tool is the investigation of its *reliability* (Haas, 1991a; Maffey-Ward et al., 1996; Atkinson and Nevill, 1998; Lachin, 2004). *Reliability* is fundamental to all aspects of clinical research, because without it we cannot have confidence in the data we collect nor can we draw rational conclusions from those data. As highlighted in Chapter 4, studies looking at cervicocephalic kinaesthetic sensibility have used different testing protocols, test repetitions and equipment to measure the parameter of interest. However, to the author's knowledge there have not been any published studies examining issues of method agreement with these varied approaches. Data regarding the level of agreement between methods would be helpful when comparing results, to see if the methods agree well enough for one method to replace another or perhaps for the two methods to be used interchangeably.

A few studies have looked at the clinical characteristics of tests for cervicocephalic kinaesthetic sensibility; namely issues of testing *reliability* (Table 28). They have all focused on the assessment of intra examiner test-retest precision although three studies also incorporated an evaluation of inter-examiner test-retest precision as well. With one exception (Kristjansson et al., 2004) the investigations were carried out solely in healthy, asymptomatic populations. A comparison of these studies was difficult as several varieties

of kinaesthetic test were utilised in these precision studies (Table 28). As highlighted in Chapter one, these can broadly be categorised into tests of repositioning to a neutral head (NHP) or straight ahead (SSA) position, repositioning to an angle within the cervical range of motion (nSA) or dynamic tasks involving head movement control. Comparison of the results from these studies was also hindered by differences in testing procedures/protocols, measuring devices, intervals between testing and types of data analysis.

One of the most significant limitations to the interpretation and comparison of results was the quality of methods/results reporting and in particular, the appropriateness and quality of statistical analyses utilised. Overall, the quality of analysis and reporting has improved since the first cervicocephalic kinaesthetic sensibility *reliability* data were published (Revel et al., 1991). This perhaps reflects that over the last 17 years an increasing awareness of statistical issues in agreement studies has been permeating the manual therapy and biomechanics research community and more statistical guidance is available in the form of accessible statistical review papers and appropriately worked examples. There has been a shift away from hypothesis tests for differences (e.g., *t*-test or one-way ANOVA) to more appropriate agreement analyses (Table 28). However, the analyses, reporting are often inadequate for the full assessment of the relative usefulness and discriminant value of the test. There is an over reliance and interpretation of the often misleading ICC statistic although in more recent studies, the level of ICC reporting has improved (e.g., type and origin of ICC) and different types of agreement estimate that complement the ICC (in the unit of measurement) have been provided to give a more comprehensive picture about the level of agreement (Table 28). Three studies utilized the Bland and Altman approach as part of the analysis (Kristjansson et al., 2001; Kristjansson et al., 2004; Strimpakos et al., 2006) although only Kristjansson (2004) appeared to provide adequate analysis outcomes. The relative paucity of *reliability* research, the differences between studies and the study limitations dictate that only tentative conclusions and inferences regarding intra- and inter-examiner test-retest precision can be drawn at this stage.

With regards to measurement equipment used in the thirty-two studies to date, seven instruments have been utilised, the most common for HRA-SSA being the laser pointer technique originally described by Revel et al., (1991). There do not appear to be any studies looking at the agreement between these instruments and hence the interchangeability of the methods.



**Table 28:** contd

Principal Author	HRA Parameter/instrument	Subject number and type	Trial Reps	Testing sessions or raters/time interval	Index of Reliability/Result
Swait, 2007	HRA-SSA & Head movement control/ <sup>c</sup> 3-Space Fastrak	16 healthy subjects	10 for HRA-SSA & 3 for tracking test	Optimum rep number for method <i>stability</i> and <i>reliability</i>  <i>Test-retest</i> 2/ same session  1/ at 1 week	<b>Stability</b>  SSA:  RMSE: Stable levels at 6 reps (mean diff, diff in SD & CV) for all movement directions  Head control test:  RMSE: Stable levels at 6 reps (mean diff, diff in SD & CV) for both tests  <b>Reliability</b>  SSA:  RMSE: Stable levels of ICC's (2,k) at 5 or more reps for all movement directions  All had lower CI for ICC >0.4  Head control test:  RMSE: Highest & most stable levels of ICC's (2,k) at 5 or more reps for both tests  Lower CI for ICC >0.4 exceeded with 1 rep.

*Abbreviations:* HRA-SSA – head repositioning accuracy to subjective straight ahead; ICC – intraclass correlation coefficient; LoA – limits of agreement; RMSE – root mean square of error (absolute error); SE – signed error; VE – variable error ; ----- studies before and after studies for this thesis.

<sup>a</sup>laser pointer technique introduced by Revel *et al* (1991) testing the ability of blindfolded subjects to accurately relocate the head, to a subjective 'straight-ahead' position, after a maximal active movement of the head in the horizontal or vertical plane.

<sup>b</sup>cervical range-of-motion device – a plastic device consisting of a magnetic yoke, resting on the shoulders, and a plastic headpiece with three goniometers positioned to measure the three cardinal planes of movement.

<sup>c</sup>3-Space Fastrak – an electromagnetic measuring instrument which tracks the positions of sensors relative to a source in three dimensions.

<sup>d</sup>Zebris CMS20/70P – a method of measurement based on the determination of the spatial coordinates around three orthogonal axes of miniature ultrasound (US) transmitters relative to a fixed system of three microphones.

The aims of the present study were to investigate:

- 1) the concurrent method agreement in HRA-SSA measurement between the laser pointer and Zebris methods of measurement.
- 2) the one-week test-retest precision (test-rest/intra-examiner *reliability*) using the laser pointer and Zebris method of measurement
- 3) the immediate inter-examiner test-retest precision (inter-examiner *reliability*) using the laser pointer and Zebris method of measurement

## **5.2 METHODS**

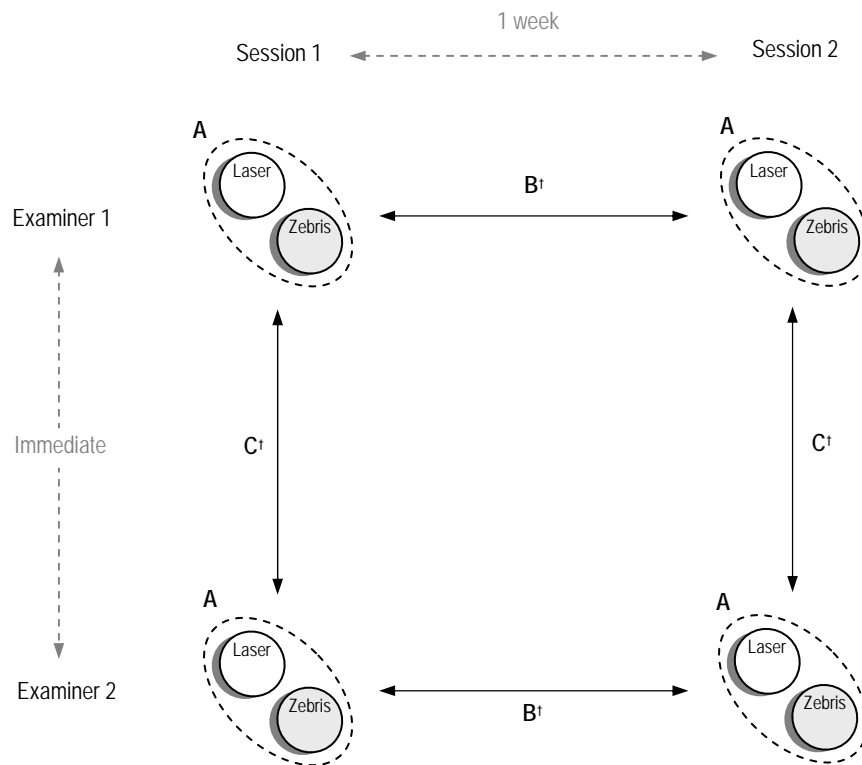
### **5.2.1 Study Setting and Design**

This study took place in the outpatient clinic at the AECC, Bournemouth, UK. Completion of questionnaires and all measurement procedures were conducted in one of the radiography rooms. The same room was used on each occasion. The study was designed to assess the method agreement of the laser pointer and Zebris methods but also the one week test-retest precision and immediate inter-examiner precision using both methods (Figure 42). As with the previous method agreement study (Chapter 4 – methods section), the present study design allowed the analysis of method agreement between the laser pointer and Zebris method of HRA measurement method with singular comparisons but also with repeated measurements. Ethical approval was obtained from the AECC research ethics committee and written consent was obtained from all subjects before entering the study.

### **5.2.2 Subject Recruitment and Selection**

The study population constituted of a convenience sample of volunteers drawn from final year AECC students. Male and female subjects between the ages of 18 to 55 years were invited to participate after an initial verbal presentation to the whole class on the nature of the investigation, risks of participation and commitment required. Volunteers were then further questioned individually by the investigator regarding their past and current medical history after verbal consent and assurance of confidentiality; if they met any of the exclusion criteria listed in Table 29, they were excluded from participation in the study.

All subjects who met the inclusion/exclusion criteria were further informed in more detail that they would need to avoid manual therapeutic procedures (e.g., massage, stretching, manipulation) applied to the thoracic and cervical spine regions during the two-week period between HRA-SSA measurements. In addition, the students were informed that during this period they would need to avoid being subjects for the purpose of teaching and practising manual procedures (palpation, manipulation set-ups, other examination techniques) involving thoracic and cervical spine regions.



**Figure 42.** Diagrammatic representation of study design. A – concurrent method agreement between laser pointer and Zebris method of HRA measurement; B – one week test-retest precision for both methods; C – immediate inter-examiner precision using both methods. † Intermediate conditions of precision with examiner, subject, time and measurement methods as the main potential sources of variability.

**Table 29.** Inclusion & Exclusion criteria

#### Inclusion

1. Age 18-55.
2. Males and females.

#### Exclusion

1. Current or previous history cervical spine pain.
2. Prior history of cervical injury or trauma.
3. Current treatment for any musculoskeletal complaint.
4. Persistent or frequent headaches (of any type).
5. Cervical Radiculopathy &/or Myelopathy
6. Inflammatory Arthritis involving C-spine
7. Tumour or infection involving C-spine
8. Vertebrobasilar Artery Insufficiency
9. Neurological disease such as MS, MN, Parkinson's, Syringomyelia etc.
10. History of Dizziness
11. Known congenital anomalies involving the C-spine.
12. Systemic disease such as Diabetes Mellitus.

#### Abbreviations:

C-Spine – Cervical Spine

MS – Multiple Sclerosis

MND – Motor Neurone Disease



Subjects were asked to report if they experienced any neck pain or discomfort during the period between measurements so their data could be omitted from the analysis. They could also discontinue participation at any time, particularly if they experienced significant discomfort before, during or after conducting the trials.

### **5.2.3 Measurement Instrumentation**

Measurement of cervicocephalic kinaesthetic sensibility was performed concurrently using the laser pointer equipment described previously (Chapter 2 – methods section) and Zebris CMS 70P system described in the last study (Chapter 4 – methods section). For this study however, the laser pointer was attached to the Zebris headpiece instead of the cycling helmet. In the previous study using the Zebris system (Chapter 4), an earphone system was developed that obliterated the pulsed sound from the transmitters but allowed a bilaterally symmetrical presentation of auditory cues and communication from the experimenter without any additional proprioceptive cues. This system also allowed the experimenter to deliver pre-recorded movement instructions (MPEG files) to the subjects. For practical reasons highlighted from the last study, this earplug system was not used. As the possible effect of auditory cues on the HRA-SSA measurements was not quantified, foam earplugs were used in the current study to obliterate the ‘ticking’ noise released by the six US transmitters. Using these earplugs, the subjects could still hear the movement instructions. The possible auditory proprioceptive cues from the examiner were standardised by delivering the pre-recorded movement instructions using small speakers positioned approximately one metre directly behind the subjects.

### **5.2.4 Measurement protocol**

Two chiropractors of seven (main investigator) and ten years clinical experience respectively were the examiners for this study. Both were experienced in assessing and treating various neuromusculoskeletal disorders and in measuring HRA in a research or clinical setting.

***Concurrent Method Agreement; Laser and Zebris method.*** Measurements took place during the daytime (between 9:00 am and 4:00 pm). As the group of subjects were generally recruited a few weeks prior to the measurement session, on arrival they were checked again by the first examiner with respect to the inclusion/exclusion criteria. They

were also given a subject information sheet to read following which they were asked to sign the consent form (Appendix IV).

For the HRA-SSA measurements the subjects were seated in a chair with a backrest for the lumbar and lower thoracic regions only. They were asked to sit as far back into the chair as possible, place the rear of their heels against specific foot markings on the floor, assume a relaxed posture and then rest forearms on their thighs. They were told that the target (laser pointer method) would be straight ahead of them but would be fitted and moved into position only after they were blindfolded. A sleeping mask was then placed over the subjects head but not so that it occluded their vision. The Zebris head and shoulder frames were then attached to the subject and checked for comfort. At this stage, the HRA-SSA testing procedure and objectives was then explained one more time.

Each subject had a similar explanation delivered in a similar manner. The SSA position was defined to the subjects as the “subjective position of the head with respect to the trunk that was vertically upright without rotation (anatomical position)”. They were also informed that on hearing the command to move the head, they should perform a maximal rotation of the head, then immediately try to relocate back to the SSA with maximum of precision and then hold this position until they received the next instruction to move. The subjects were also instructed to avoid movement of the trunk during these repositioning tasks. No speed instruction was given. No feedback on performance was given during the testing session

The Zebris system was then turned on and each subject were then asked to fit the foam earplugs to obliterate the sound of the US transmitters but so that they could clearly still hear the movement instructions from the speakers behind them. New earplugs were used with each subject. Having completed this, all subjects then had their vision occluded using the sleeping mask for the remainder of the testing procedure. All were instructed to keep their eyes closed behind the mask. The target for the laser pointer method was then fitted and moved to 90 cm in front of the subject (front of the laser pointer) and the laser pointer switched on. After the target was positioned, the room lights were switched off.

The test procedure was divided into two parts: 1) determining the SSA and 2) performing the HRA-SSA task. The subject was first asked to find what they perceived as the ‘straight ahead’ (SSA) position of their head with respect to the trunk as previously defined and

notify the examiner verbally when they felt correctly aligned. The target was then quickly moved so that the laser pointer's light beam projected to roughly the centre of the target and cross was made on the target to represent the  $0^\circ$  position. The Zebris system was then set to record mode, 'zeroed' (calibrated) so that this position was defined as  $0^\circ$  for this equipment. The subject was then asked to recognise and remember this position for 2-3 seconds and then reminded that this was the SSA position or target that they were to try to relocate back to as accurately as possible after each movement in the trial. To initiate the repositioning task (as described earlier), the subject was then simply instructed to move the head in the direction requested using the pre-recorded movement instructions (MPEG files). Once the subject had appeared to stop moving at the reposition point, the point where the light beam stopped on the target was marked with a dot from a pen and labelled according to the repetition number. The examiner waited approximately 6-8 seconds before issuing the next instruction to move. Five repetitions of HRA to the SSA position were undertaken with each trial direction. Four repositioning trials were performed by each subject; two around the  $\theta Y$  axis of rotation (LR & RR  $\Rightarrow$  0) and two around the  $\theta X$  axis of rotation (Ext & Flex  $\Rightarrow$  0). A new SSA position was established for each trial after approximately two minutes rest between each trial. The sequence of trials was varied between subjects to minimize the possible effect of order on between trial comparisons. No feedback on performance was given either during or at the end of a testing session. At the end of the testing session, the equipment was removed from the subjects. To further minimise subject feedback, the blindfold was removed only after the target for the laser pointer had been moved away.

***Immediate Inter-Examiner Precision.*** All subjects were tested again at the same session by a second examiner after a five minute break. The second examiner performed the entire set-up and measurement protocol and was blind to the previous results. For each subject, the same sequence of HRA trials was performed by both examiners. The order of examiners was alternated between subjects to minimize the possible effect of order on between examiner comparisons.

***One-week Test-Retest Precision.*** The subjects were retested by both examiners after a one week interval. Where possible, subjects were examined at the same time of the day. The same sequence of HRA trials and examiner order was used at the retest session. At each retest session, the subjects were also checked by the first examiner to see whether they had experienced any neck related complaints or received any manual procedures to the cervico-

thoracic region since the last session, and whether there were any other factors that may affect kinaesthetic performance.

### **5.2.5 Data Processing**

The raw angular 3-D position data for each trial, processed and saved by the Zebris system's dedicated data acquisition software (WinData), was first exported to a storage folder as an ASCII file. Each file name was then recoded by the PhD supervisor (Prof. J. Bagust) to ensure the examiners were blinded to the subject and session when analysing the raw data. The data in this format was then imported into LabVIEW for measurement analysis as described in the previous study (Chapter 4 – methods section). All Zebris data was analysed by examiner 1 (the main study investigator). The raw data obtained from the laser pointer method were processed as described in chapter 2 (methods section). Each examiner independently processed the results they had obtained after the paper target sheets were again recoded by the PhD supervisor to ensure blinding. After the both sets of raw data were processed, the data sets were then decoded to reveal subject, session and examiner. For each subject, the data used for descriptive and inferential statistical analysis consisted of the mean value of the repetitions in each trial. These values were calculated using Excel 2000 database software (Microsoft<sup>®</sup> Corp., Redmond, WA, USA).

### **5.2.6 Data Analysis**

The absolute (unsigned) error (AE) and the signed error (SE) were used for comparative analysis of HRA-SSA for each study. The AE was derived from the original directional data (SE) as described in the previous study (Chapter 4 – methods section). All data sets (AE and SE) were examined for normality using a combination of statistical testing (Kolmogorov-Smirnov or K-S test) and observation of the data using frequency histograms, Q-Q (normal) plots and box plot summaries of the measures of central tendency (mean and median) and distribution.

Analyses of concurrent method agreement between the laser pointer and Zebris method of HRA measurement (Figure 42A), one week test-retest precision for both methods (Figure 42B) and immediate inter-examiner precision using both methods (Figure 42C) were conducted using the Bland and Altman approach (Altman and Bland, 1983; Bland and Altman, 1986; Bland and Altman, 1999) again as described in the previous study (Chapter 4

– methods section & Appendix III). Decisions on how to condense the HRA data were made post hoc after analysis of all the direction error components were analysed. All calculations and distribution plots for the Bland and Altman analysis were produced using Excel 2000 database software. All other statistical analyses related to this approach (e.g., statistical and graphical checking of assumptions and regression analysis) were carried out using SPSS for Windows.

## 5.3 RESULTS

### 5.3.1 Group demographics and characteristics

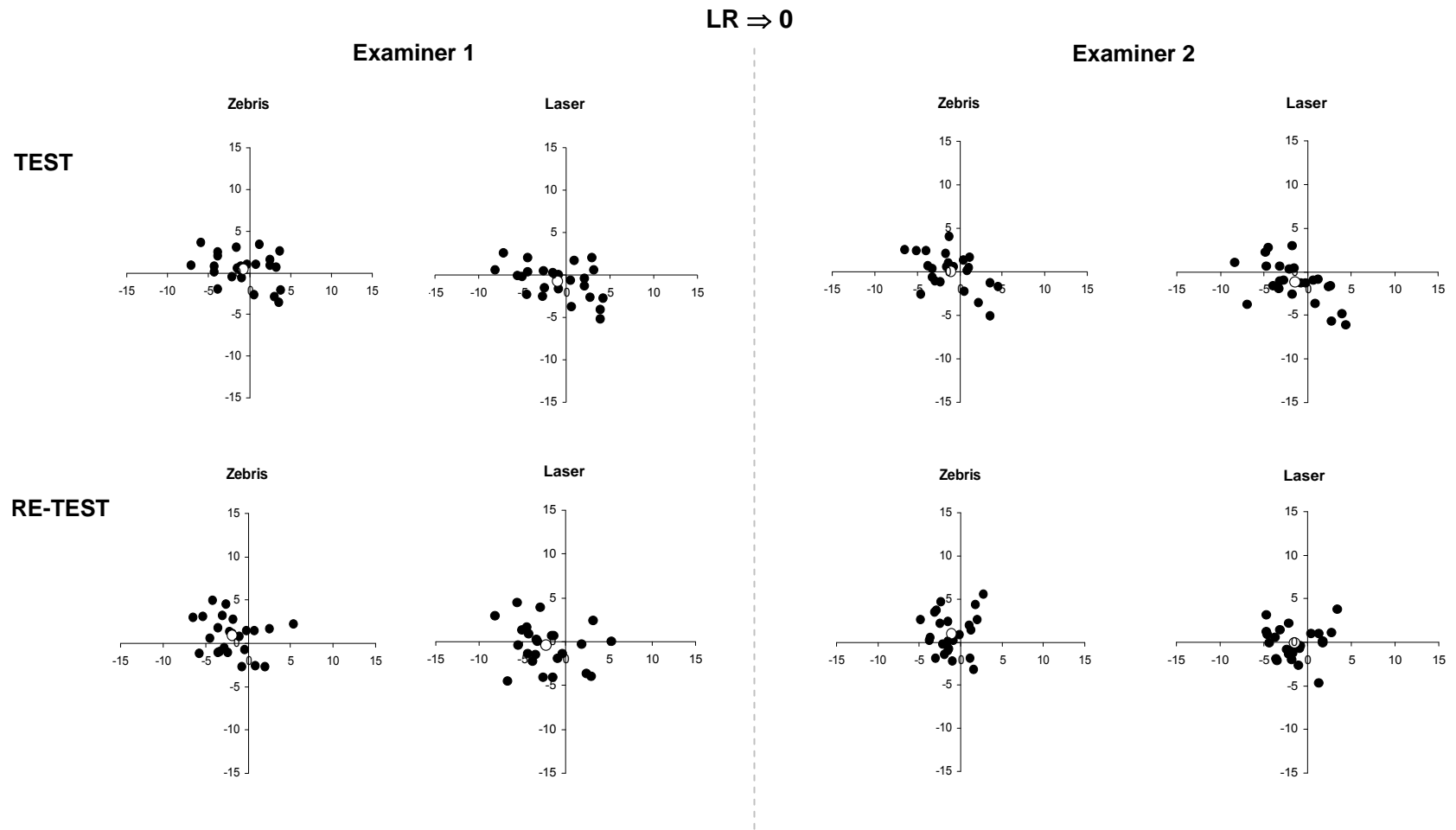
Twenty-four healthy subjects (12 men, 12 women) aged 22-38 years (mean age  $\pm$  SD = 26.0  $\pm$  3.6 yrs) agreed and were eligible to participate in the investigation after the initial verbal presentation. Every one of the subjects performed the necessary tests over the one-week study period and all data sets were complete. No adverse effects from the testing were reported either during or immediately after the study.

### 5.3.2 Kinaesthetic Sensibility Testing

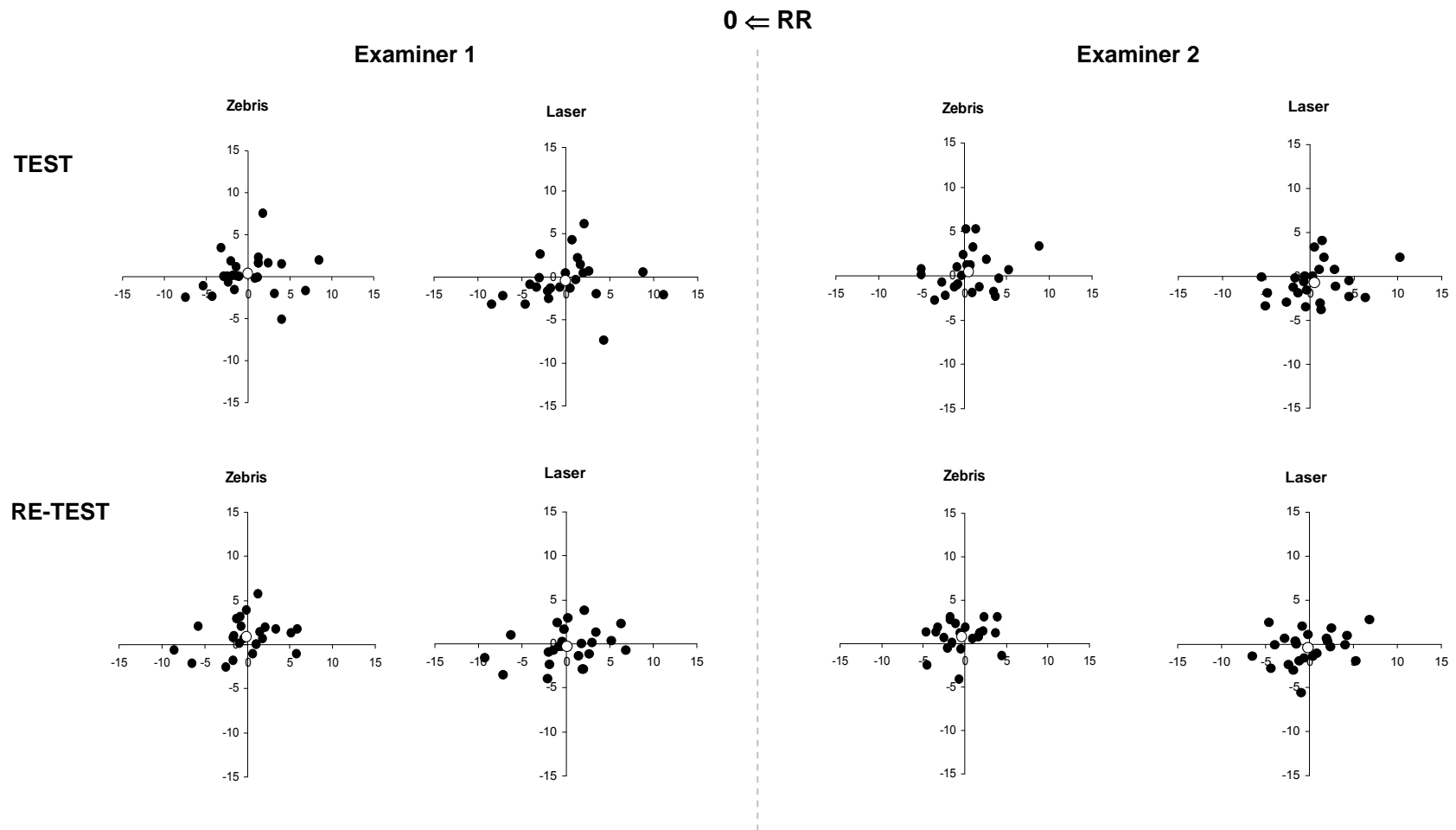
An overview of the individual subject HRA-SSA values (mean values of the five repositioning repetitions) for each of the trials at the test and retest sessions and with each examiner, is presented in Figures 43 and 46 using scatter plots. A comparison of the distribution and measures of central tendency for the absolute (unsigned) and signed error for the group at each trial is presented in Appendix IV using box plots. All data sets passed the normality testing using the Kolmogorov-Smirnov test ( $P > 0.1$ ) and were generally considered to exhibit a normal distribution after observation of the graphical summaries. As such, the descriptive results for the HRA-SSA were summarized using the mean, standard deviation (SD) and 95% confidence intervals for the mean (95% CI) (Tables 30 and 31).

Using the Zebris system the mean absolute HRA-SSA for the horizontal rotation trials (L and R  $\Rightarrow$  0), ranged from 2.1° to 3.0° and 1.6° to 2.0° for the  $\theta Y$  and  $\theta X$  error components respectively. The mean absolute HRA-SSA for the vertical rotation trials (Ext and Flex  $\Rightarrow$  0) ranged from 0.9° to 1.8° and 2.0° to 3.1° for the  $\theta Y$  and  $\theta X$  error components.

Using the laser pointer system the mean ( $\pm$  SD) absolute HRA-SSA for the horizontal rotation trials (L & R  $\Rightarrow$  0), ranged from 2.5° to 3.5° and 1.5° to 2.2° for the  $\theta Y$  and  $\theta X$  error components respectively. The mean absolute HRA-SSA for the vertical rotation trials (Ext and Flex  $\Rightarrow$  0) ranged from 1.1° to 2.0° and 2.5° to 3.6° for the  $\theta Y$  and  $\theta X$  error components. The absolute repositioning errors were greatest in the direction of primary motion for all the trials. The signed repositioning errors highlighted a tendency for subjects to undershoot the target with both the horizontal and vertical repositioning trials. For the Bland and Altman analyses, the differences between methods, testing sessions and examiners were also considered normal in distribution for both measuring devices.

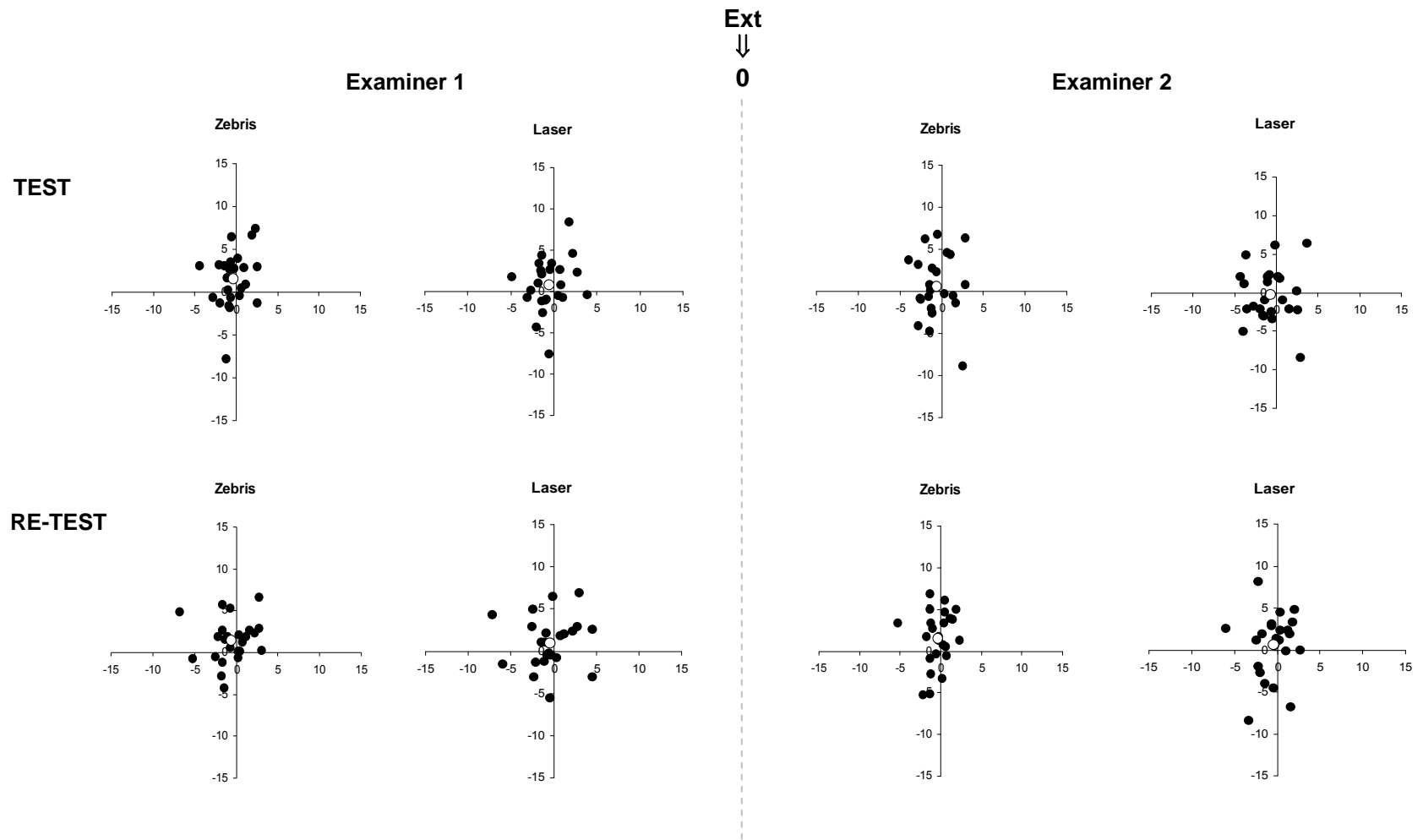


**Figure 43.** Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. All data are presented as degrees. Abbreviations; left rotation (LR); the empty circle (O) on the scatter plots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).

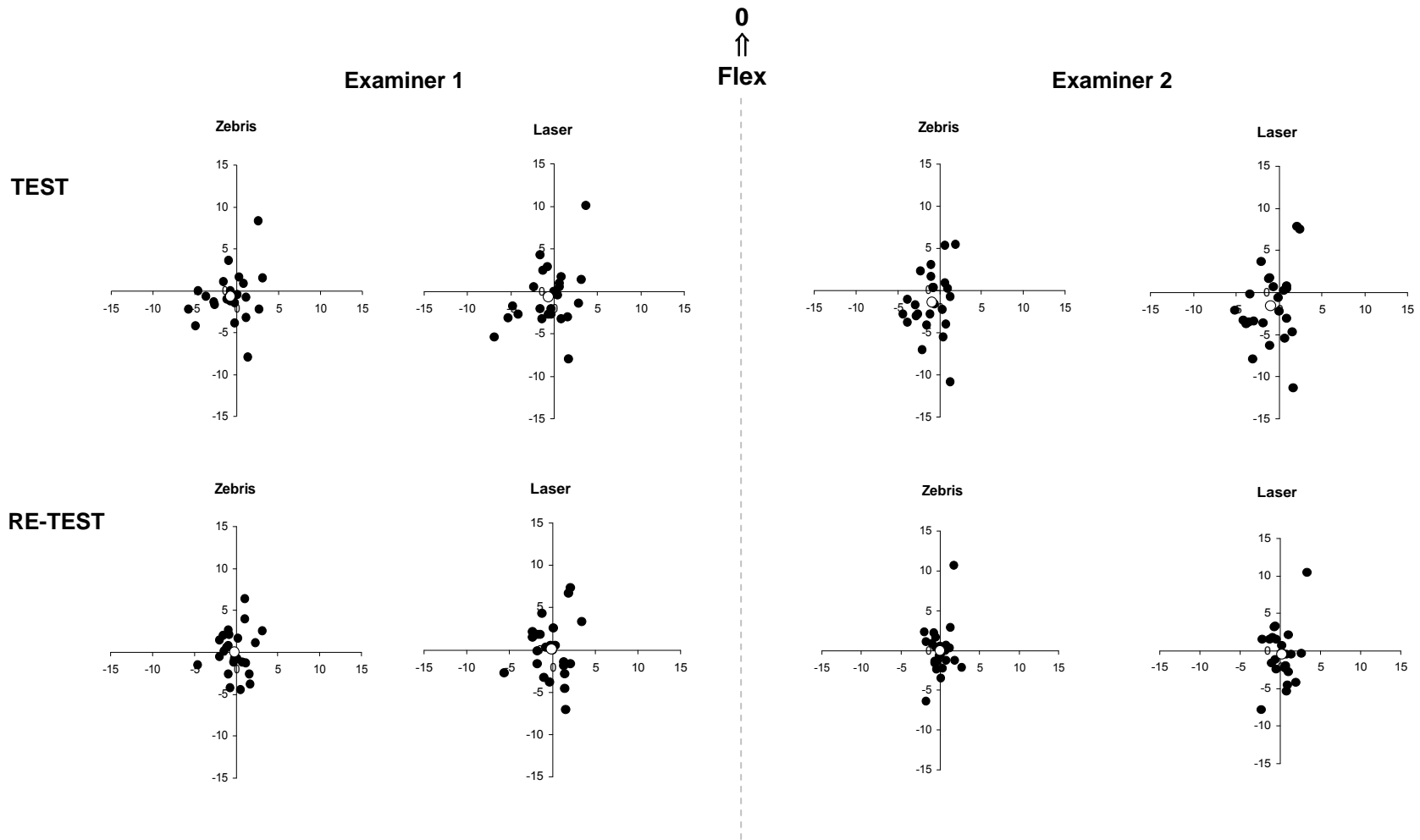


**Figure 44.** Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. All data are presented as degrees. Abbreviations; right rotation (RR); the empty circle (O) on the scatter plots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).





**Figure 45.** Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. All data are presented as degrees. Abbreviations; Extension (Ext); the empty circle (O) on the scatter plots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).



**Figure 46.** Scatter plots showing the individual subject HRA-SSA as measured by Examiners 1 & 2 using the laser pointer method and Zebris at the test and retest sessions. All data are presented as degrees. Abbreviations; flexion (Flex); the empty circle (O) on the scatter plots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).

**Table 30.** Absolute (unsigned) vertical ( $\theta X$ ), horizontal ( $\theta Y$ ) repositioning error (degrees) in healthy subjects (n =24) measured concurrently using the laser pointer and Zebris CMS70P method in the horizontal and vertical movement plane by two examiners at sessions two weeks apart.

Primary Motion	Axis of rotation	Examiner 1								Examiner 2							
		Test				Retest				Test				Retest			
		Laser		Zebris		Laser		Zebris		Laser		Zebris		Laser		Zebris	
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI
LR	$\theta X$	1.7 $\pm$ 1.41	1.07 $\rightarrow$ 2.25	1.6 $\pm$ 1.26	1.16 $\rightarrow$ 2.13	2.0 $\pm$ 1.55	1.37 $\rightarrow$ 2.68	2.0 $\pm$ 1.21	1.48 $\rightarrow$ 2.50	2.2 $\pm$ 1.64	1.44 $\rightarrow$ 2.82	1.7 $\pm$ 1.31	1.10 $\rightarrow$ 2.21	1.5 $\pm$ 1.16	0.97 $\rightarrow$ 1.95	2.0 $\pm$ 1.51	1.40 $\rightarrow$ 2.68
	$\theta Y^\dagger$	3.2 $\pm$ 1.26	1.04 $\rightarrow$ 2.56	2.8 $\pm$ 1.74	2.04 $\rightarrow$ 3.51	3.5 $\pm$ 1.90	2.73 $\rightarrow$ 4.33	2.8 $\pm$ 1.78	2.02 $\rightarrow$ 3.52	3.0 $\pm$ 1.96	2.20 $\rightarrow$ 3.85	2.5 $\pm$ 1.65	1.85 $\rightarrow$ 3.24	2.5 $\pm$ 1.30	1.92 $\rightarrow$ 3.01	2.1 $\pm$ 1.09	1.62 $\rightarrow$ 2.54
RR	$\theta X$	2.1 $\pm$ 1.78	1.32 $\rightarrow$ 2.83	1.8 $\pm$ 1.71	1.03 $\rightarrow$ 2.48	1.6 $\pm$ 1.24	1.11 $\rightarrow$ 2.16	1.7 $\pm$ 1.31	1.12 $\rightarrow$ 2.23	1.8 $\pm$ 1.29	1.30 $\rightarrow$ 2.39	1.8 $\pm$ 1.41	1.20 $\rightarrow$ 2.39	1.5 $\pm$ 1.27	1.00 $\rightarrow$ 2.06	1.6 $\pm$ 1.02	1.16 $\rightarrow$ 2.02
	$\theta Y^\dagger$	3.3 $\pm$ 2.87	2.12 $\rightarrow$ 4.54	3.0 $\pm$ 2.13	2.13 $\rightarrow$ 3.92	2.9 $\pm$ 2.55	1.83 $\rightarrow$ 4.00	2.6 $\pm$ 2.36	1.59 $\rightarrow$ 3.58	2.7 $\pm$ 2.45	1.62 $\rightarrow$ 3.69	2.4 $\pm$ 2.14	1.44 $\rightarrow$ 3.25	2.7 $\pm$ 1.92	1.84 $\rightarrow$ 3.47	2.2 $\pm$ 1.35	1.64 $\rightarrow$ 2.78
Ext	$\theta X^\dagger$	2.5 $\pm$ 2.15	1.58 $\rightarrow$ 3.39	2.8 $\pm$ 2.24	1.84 $\rightarrow$ 3.73	2.5 $\pm$ 1.90	1.67 $\rightarrow$ 3.27	2.3 $\pm$ 1.85	1.48 $\rightarrow$ 3.03	2.8 $\pm$ 1.96	2.00 $\rightarrow$ 3.66	3.0 $\pm$ 2.37	1.97 $\rightarrow$ 3.97	3.0 $\pm$ 2.30	1.99 $\rightarrow$ 3.93	3.0 $\pm$ 1.90	2.23 $\rightarrow$ 3.84
	$\theta Y$	1.7 $\pm$ 1.13	1.19 $\rightarrow$ 2.14	1.4 $\pm$ 1.00	0.97 $\rightarrow$ 1.82	2.0 $\pm$ 1.89	1.22 $\rightarrow$ 2.82	1.8 $\pm$ 1.54	1.19 $\rightarrow$ 2.49	2.0 $\pm$ 1.33	1.40 $\rightarrow$ 2.52	1.7 $\pm$ 0.93	1.34 $\rightarrow$ 2.12	1.5 $\pm$ 1.32	0.95 $\rightarrow$ 2.06	1.2 $\pm$ 1.06	0.75 $\rightarrow$ 1.65
Flex	$\theta X^\dagger$	2.8 $\pm$ 2.33	1.80 $\rightarrow$ 3.76	2.2 $\pm$ 2.17	1.23 $\rightarrow$ 3.06	2.7 $\pm$ 2.10	1.77 $\rightarrow$ 3.55	2.1 $\pm$ 1.53	1.45 $\rightarrow$ 2.74	3.6 $\pm$ 2.91	2.35 $\rightarrow$ 4.81	3.1 $\pm$ 2.48	2.02 $\rightarrow$ 4.12	2.7 $\pm$ 2.42	1.67 $\rightarrow$ 3.71	2.0 $\pm$ 2.27	1.04 $\rightarrow$ 2.96
	$\theta Y$	2.0 $\pm$ 1.80	1.22 $\rightarrow$ 2.74	1.8 $\pm$ 1.61	1.17 $\rightarrow$ 2.52	1.6 $\pm$ 1.19	1.08 $\rightarrow$ 2.08	1.3 $\pm$ 1.02	0.86 $\rightarrow$ 1.72	1.9 $\pm$ 1.36	1.33 $\rightarrow$ 2.48	1.6 $\pm$ 1.16	1.14 $\rightarrow$ 2.12	1.1 $\pm$ 0.81	0.80 $\rightarrow$ 1.48	0.9 $\pm$ 0.72	0.64 $\rightarrow$ 1.25

Abbreviations: left and right rotation (LR, RR), extension (Ext), forward flexion (Flex).

$\dagger$  Primary axis of rotation for the repositioning test

**Table 31.** Signed repositioning error (degrees) in healthy subjects (n =24) measured concurrently using the laser pointer and Zebris CMS70P method in the horizontal and vertical movement plane by two examiners at sessions two weeks apart. These data represent the overshoot/undershoot characteristics in the plane of movement for HRA-SSA.

		Examiner 1								Examiner 2							
		Test				Retest				Test				Retest			
		Laser		Zebris		Laser		Zebris		Laser		Zebris		Laser		Zebris	
Primary Motion	Axis of rotation	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI	Mean ± SD	95%CI
LR	θY <sup>†</sup>	-0.9 ± 3.71	-2.50 → 0.64	-0.7 ± 3.26	-2.04 → 0.71	-2.2 ± 3.40	-3.63 → - -0.76	-1.78 ± 2.80	-2.97 → -0.60	-1.4 ± 3.37	-2.83 → 0.01	-1.0 ± 2.90	-2.32 → 0.22	-1.4 ± 2.45	-2.43 → -0.37	-1.1 ± 2.11	-1.98 → -0.20
RR	θY <sup>†</sup>	0.0 ± 4.45	-1.84 → 1.92	0.1 ± 3.75	-1.53 → 1.63	0.2 ± 3.92	-1.44 → 1.87	-0.1 ± 3.54	-1.58 → 1.41	0.6 ± 3.60	-0.92 → 2.12	0.5 ± 3.16	-0.79 → 1.88	-0.1 ± 3.32	-1.48 → 1.32	-0.3 ± 2.61	-1.39 → 0.82
Ext	θX <sup>†</sup>	0.8 ± 3.23	-0.58 → 2.15	1.5 ± 3.29	0.10 → 2.87	0.9 ± 3.01	-0.34 → 2.20	1.4 ± 2.58	0.31 → 2.49	-0.3 ± 3.48	-1.79 → 1.15	0.5 ± 3.81	-1.11 → 2.11	0.6 ± 3.76	-1.02 → 2.15	1.5 ± 3.31	0.09 → 2.88
Flex	θX <sup>†</sup>	-0.7 ± 3.59	-2.26 → 0.78	-0.7 ± 2.99	-2.00 → 0.53	0.1 ± 3.44	-1.40 → 1.51	0.0 ± 2.63	-1.11 → 1.11	-1.6 ± 4.40	-3.41 → 0.30	-1.5 ± 3.71	-3.02 → 0.11	-0.5 ± 3.62	-2.04 → 1.02	-0.1 ± 3.06	-1.37 → 1.22

Abbreviations: left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); HRA-SSA – head repositioning accuracy to subjective straight ahead.

<sup>†</sup> Primary axis of rotation for the repositioning test

### 1) Concurrent Method Agreement for HRA-SSA; Laser and Zebris method.

A visual examination of the distribution plots followed by a statistical evaluation the relationship between the differences and the mean values (Pearson's  $r$  and absolute residual analysis), indicated the presence of a *proportional bias* and/or *heteroscedasticity* for some of the method agreement (laser vs Zebris) HRA-SSA directional components. For comparative purposes the Bland and Altman results for these data sets, are initially presented using the standard analysis approach (non-regression adjusted).

#### *Concurrent method agreement (laser vs Zebris): absolute values.*

The standard method agreement results using a repeated measurements Bland and Altman analysis are shown in Table 32 for examiner one (principal investigator) and in Table 33 for the second examiner. Overall, the standard and regression adjusted results indicate that there was no statistically significant *fixed bias* ( $d$ ) between the two methods of measurement with either examiner. For all measurements, the estimates of *fixed bias* ( $d$ ) were less than 0.6 degrees. Estimates of *random error* with the standard (unadjusted) method agreement analysis were presented using both the repeated measures variance ( $\pm 1.96s_c$ ) and single session measurements taken at the retest session ( $\pm 1.96s_d$ ). This was mainly to allow a comparison with data for which regression adjusted agreement estimates were calculated. The regression analysis does not incorporate the same adjustment of variance ( $s_d$  to  $s_c$ ) as with the standard replicate measures approach. It also allowed an immediate assessment of the effects of test-retest precision on method agreement. Using the standard Bland and Altman analysis with repeated measures, estimates of *random error* ( $\pm 1.96s_c$ ) appeared similar in magnitude between examiners (Tables 32 and 33). For the Flex and Ext  $\Rightarrow 0$  trials, the errors were relatively larger around the axis of primary of motion ( $\theta X$ ). Estimates of *random error* ( $\pm 1.96s_c$ ) around the axis of primary of motion for the L and RR  $\Rightarrow 0$  trials ( $\theta Y$ ) ranged from 2.5 to 3.6 degrees and from 3.4 to 3.8 degrees for the axis of primary of motion with the Flex and Ext  $\Rightarrow 0$  trials ( $\theta X$ ). The primary motion axis results from the single session measurements ( $\pm 1.96s_d$ ) ranged from 1.2 to 2.7 for the L and RR  $\Rightarrow 0$  trials ( $\theta Y$ ) and 1.7 to 2.8 with the Flex and Ext  $\Rightarrow 0$  trials ( $\theta X$ ). These unadjusted results suggest an increase in the estimate of method agreement *random error* when the variances from test-retest measurements are incorporated in the analysis.

**Table 32.** Results of Bland and Altman analysis for method agreement using repeated measurements by examiner 1 between two different measurement methods (Zebris and Laser) for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Rotation	d	1.96s <sub>c</sub>	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
							95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.03	3.45	2.92	0.36	-0.77 → 0.72	-3.49 → 3.43	-4.41 → -2.56	2.51 → 4.35
	$\theta Y^\dagger$	-0.60	3.23	1.16	0.34	-1.30 → 0.10	-3.83 → 2.64	-4.74 → -2.93	1.73 → 3.54
RR	$\theta X$	-0.14	3.04	2.51	0.32	-0.80 → 0.52	-3.19 → 2.91	-3.99 → -2.38	2.11 → 3.71
	$^*\theta Y^\dagger$	-0.32	3.04	1.23	0.32	-0.97 → 0.33	-3.35 → 2.71	-4.15 → -2.55	1.91 → 3.51
Ext	$\theta X^\dagger$	0.04	3.57	1.74	0.37	-0.72 → 0.81	-3.51 → 3.60	-4.45 → -2.57	2.66 → 4.54
	$\theta Y$	-0.24	2.33	1.49	0.24	-0.74 → 0.27	-2.56 → 2.09	-3.18 → -1.94	1.47 → 2.71
Flex	$^*\theta X^\dagger$	-0.60	3.45	2.02	0.36	-1.35 → 0.14	-4.06 → 2.86	-4.97 → -3.15	1.95 → 3.77
	$\theta Y$	-0.22	2.74	1.51	0.29	-0.80 → 0.38	-2.96 → 2.54	-3.71 → -2.21	1.79 → 3.29

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement

d is the mean of differences (Zebris – Laser means), +ve value indicates Zebris > Laser; s<sub>c</sub> is the adjusted SD of the differences; s<sub>d</sub> is the SD of the differences without adjusted variation (retest session)

\* Data sets (replicate) exhibiting proportional bias &/or heteroscedasticity.

† Primary axis of rotation for the repositioning test.

**Table 33.** Results of Bland and Altman analysis for method agreement using repeated measurements by Examiner 2 between two different measurement methods (Zebris and Laser) for the absolute X & Y-axis head repositioning accuracy (degrees).

Primary Motion	Axis of Rotation	d	1.96s <sub>c</sub>	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
							95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	0.04	3.33	2.92	0.35	-0.67 → 0.76	-3.28 → 3.37	-4.17 → -2.40	2.49 → 4.26
	$^*\theta Y^\dagger$	-0.44	2.47	1.29	0.26	-0.97 → 0.09	-2.90 → 2.02	-3.55 → -2.25	1.37 → 2.67
RR	$\theta X$	0.01	3.12	2.86	0.32	-0.66 → 0.68	-3.10 → 3.11	-3.95 → -2.26	2.27 → 3.96
	$^*\theta Y^\dagger$	-0.38	3.64	2.65	0.38	-1.17 → 0.41	-4.03 → 3.27	-5.00 → -3.07	2.31 → 4.24
Ext	$\theta X^\dagger$	0.11	3.35	2.78	0.35	-0.61 → 0.83	-3.24 → 3.46	-4.13 → -2.36	2.57 → 4.34
	$^*\theta Y$	-0.27	2.10	1.10	0.22	-0.72 → 0.18	-2.36 → 1.82	-2.94 → -1.79	1.25 → 2.39
Flex	$^*\theta X^\dagger$	-0.59	3.80	2.21	0.40	-1.41 → 0.23	-4.40 → 3.21	-5.40 → -3.39	2.21 → 4.21
	$\theta Y$	-0.23	2.14	1.12	0.22	-0.69 → 0.23	-2.36 → 1.90	-2.94 → -1.78	1.32 → 2.48

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement

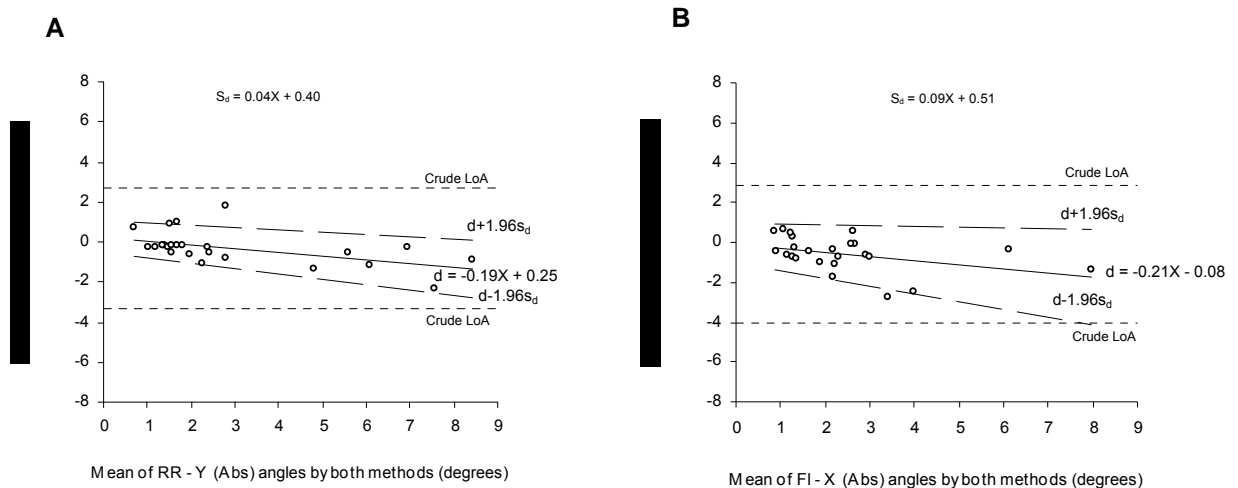
d is the mean of differences (Zebris - Laser), +ve value indicates Zebris > Laser; s<sub>c</sub> is the adjusted SD of the differences; s<sub>d</sub> is the SD of the differences without adjusted variation (retest session)

\* Data sets (replicate) exhibiting proportional bias &/or heteroscedasticity.

† Primary axis of rotation for the repositioning test.

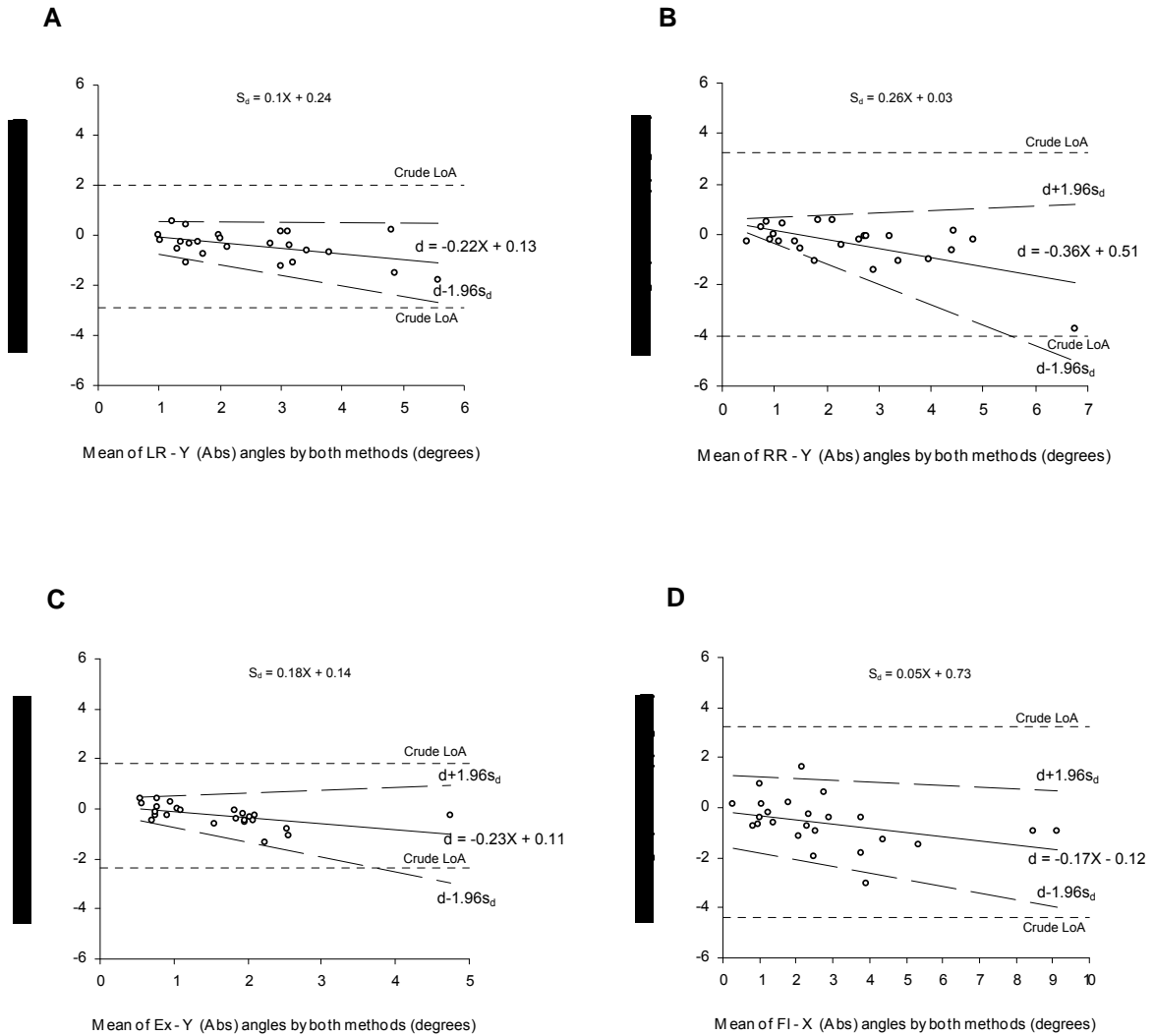
When the data were examined for relationships between the differences and the means, a small but statistically significant *proportional bias* was seen with examiner one for the RR  $\Rightarrow$  0  $\theta$ Y and Flex  $\Rightarrow$  0  $\theta$ X error components (Figure 47). A small but statistically insignificant *heteroscedastic* relationship was also apparent. With measurements obtained by examiner two, statistically significant *proportional bias* and *heteroscedastic* relationships were exhibited with the  $\theta$ Y error components for the LR, RR and Ext  $\Rightarrow$  0 trials and the  $\theta$ X error components for the Flex  $\Rightarrow$  0 trial (Figure 48). The adjusted bias and agreement estimates (regression analysis) for the affected data sets are shown in Table 34. All *proportional biases* were observed in the same negative direction and less than 0.4; the majority were 0.2 (Table 34). This represents proportionally larger measurements with the laser method compared with the Zebris with increasing mean measurement error. As previously mentioned, the *fixed bias* estimates were not statistically affected by the *proportional bias*. When *proportional bias* and *heteroscedasticity* were present a comparison of the adjusted results with the unadjusted single session data revealed a progressive shift the LoA's towards relatively larger laser pointer measurements with higher mean error scores and a relative narrowing of the LoA for low mean values and widening for higher values.

### Examiner 1



**Figure 47.** Plots of the absolute difference against mean (repeated measures) between Zebris and Laser methods by Examiner 1 for the head repositioning accuracy (HRA). On each plot; systematic error or bias is indicated by the mean value line (d); 'limits of agreement' (LoA) are given by the  $d \pm 1.96s_d$  lines. A - right rotation Y-axis; B - Flexion X-axis.

Examiner 2



**Figure 48.** Plots of the absolute difference against mean (repeated measures) between Zebris and Laser methods by Examiner 2 for the head repositioning accuracy (HRA). On each plot; systematic error or bias is indicated by the mean value line (d); 'limits of agreement' (LoA) are given by the  $d \pm 1.96s_d$  lines. A - right rotation Y-axis; B - Flexion X-axis.



**Table 34.** Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=24) between the two different measurement methods (Zebris and Laser) with data sets exhibiting a relationship between the mean errors and differences.

Primary Motion	Examiner	Axis of Motion	Fixed Bias	95% CI for d	Prop bias	Lowest mean HRA value		Highest mean HRA value	
						d ± 1.96s <sub>d</sub>	95% limits of agreement	d ± 1.96s <sub>d</sub>	95% limits of agreement
LR	2	θY	0.1	-0.36 → 0.61	-0.2	-0.1 ± 0.7	-0.8 → 0.6	-1.1 ± 1.6	-2.7 → 0.5
	1	θY	0.3	-0.36 → 0.76	-0.2	0.1 ± 0.8	-0.7 → 0.9	-1.3 ± 1.4	-2.7 → 0.1
RR	2	θY	0.5	-0.07 → 1.08	-0.4	0.3 ± 0.3	0.0 → 0.6	-1.9 ± 3.5	-5.4 → 1.6
	2	θY	0.1	-0.21 → 0.42	-0.2	0.0 ± 0.5	-0.5 → 0.5	-1.0 ± 2.0	-3.0 → 1.0
Flex	1	θX	-0.1	-0.71 → 0.54	-0.2	-0.3 ± 1.2	-1.5 → 0.9	-1.8 ± 2.4	-4.2 → 0.6
	2	θX	-0.1	-0.76 → 0.53	-0.2	-0.2 ± 1.5	-1.7 → 1.3	-1.7 ± 2.3	-4.0 → 0.6

*Abbreviations:* left and right rotation (LR, RR); Prop - proportional  
d is the mean of differences (Zebris - Laser means) ; a +ve value indicates Zebris > Laser.

**Concurrent method agreement (laser vs Zebris): signed values.** The standard Bland and Altman method agreement analysis results using a repeated measurements are shown in Table 35 for examiner one and Table 36 for the second examiner. Overall, the standard and regression adjusted results indicate that there was no statistically significant *fixed bias* (d) between the two methods of measurement with either examiner. For all measurements, the estimates of *fixed bias* (d) were less than 0.9 degrees; with regression adjustment the majority (6 of 8) were 0.1 degrees (Table 37). Estimates of *random error* with standard (unadjusted) the method agreement analysis are again presented using both the repeated measures variance ( $\pm 1.96s_c$ ) and single session measurements taken at the retest session ( $\pm 1.96s_d$ ). Using the standard Bland and Altman analysis with repeated measures, the unadjusted estimates of *random error* ( $\pm 1.96s_c$ ) ranged from 3.3 to 4.7 degrees. They appeared similar in magnitude between examiners and between trials (Table 35 and Table 36). The results from the single session measurements ( $\pm 1.96s_d$ ) were noticeably smaller ranging from 1.2 to 2.3 degrees. As with the absolute (unsigned) measurements, these unadjusted results suggest an increase in the estimate of method agreement *random error* when the variances from test-retest measurements are incorporated in the analysis.

**Table 35.** Results of Bland and Altman analysis for method agreement using repeated measurements by Examiner 1 between two different measurement methods (Zebris and Laser) for the signed head repositioning accuracy (degrees) in the primary motion direction

Primary Motion	Axis of Rotation	d	1.96s <sub>c</sub>	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
							95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	*θY	0.34	4.60	1.45	0.48	-0.65 → 1.33	-4.26 → 4.94	-5.52 → -2.99	3.67 → 6.20
RR	*θY	-0.14	4.39	1.67	0.46	-1.08 → 0.81	-4.53 → 4.26	-5.73 → -3.33	3.06 → 5.45
Ext	θX	0.59	4.25	1.94	0.44	-0.33 → 1.50	-3.66 → 4.83	-4.79 → -2.53	3.70 → 5.96
Flex	*θX	-0.03	4.15	2.25	0.43	-0.92 → 0.87	-4.19 → 4.13	-5.28 → -3.10	3.04 → 5.23

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Zebris [mean HRA] minus Laser [mean HRA]); s<sub>c</sub> is the adjusted SD of the differences; s<sub>d</sub> is the SD of the differences without adjusted variation

\* Data sets exhibiting proportional bias &/or heteroscedasticity.

**Table 36.** Results of Bland and Altman analysis for method agreement using repeated measurements by examiner 1 between two different measurement methods (Zebris and Laser) for the signed head repositioning accuracy (degrees) in the primary motion direction

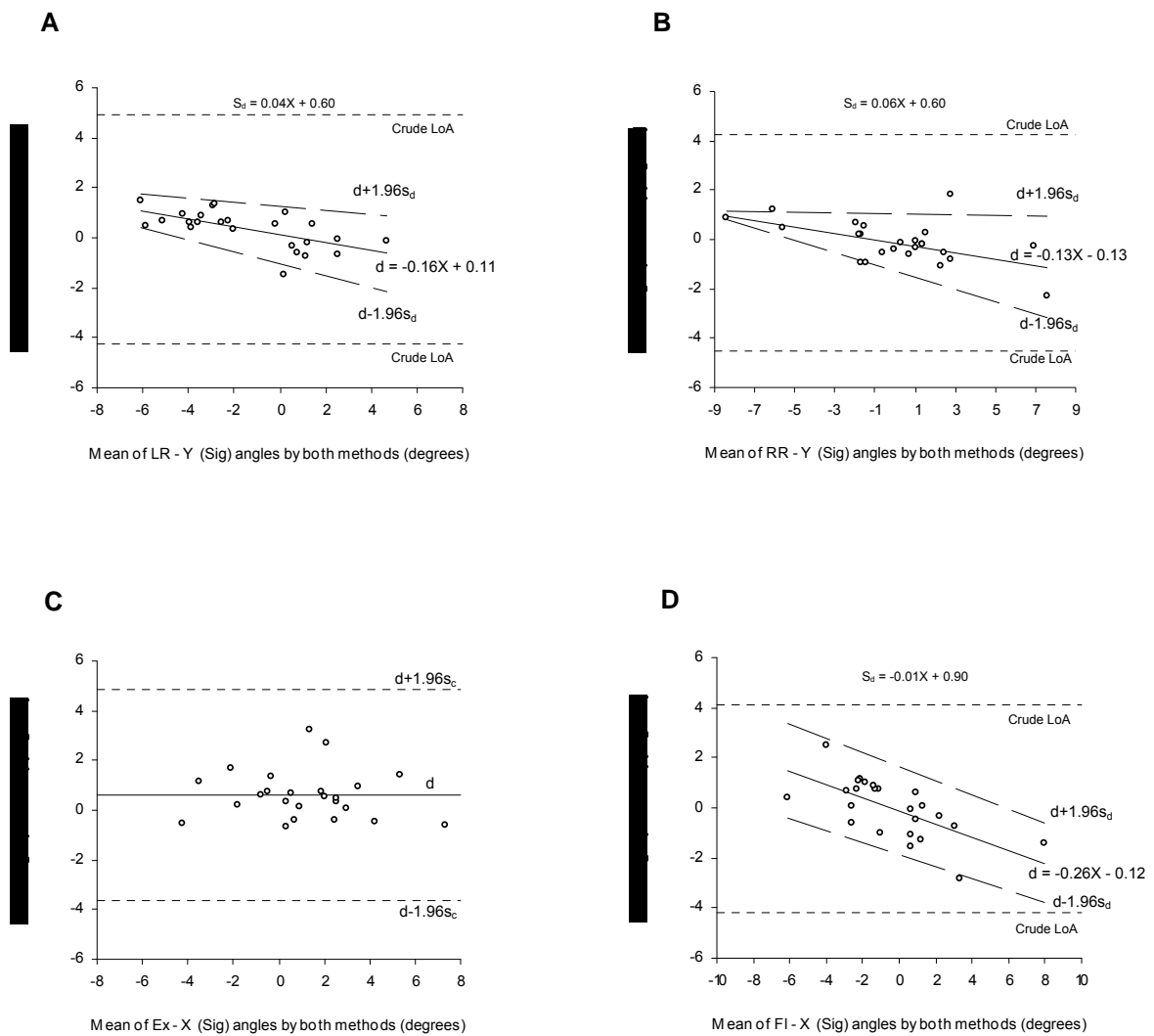
Primary Motion	Axis of Rotation	d	1.96s <sub>c</sub>	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
							95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	*θY	0.35	3.27	1.23	0.34	-0.36 → 1.05	-2.92 → 3.62	-3.81 → -2.04	2.74 → 4.50
RR	*θY	-0.13	4.51	1.92	0.47	-1.10 → 0.84	-4.64 → 4.37	-5.84 → -3.43	3.17 → 5.57
Ext	θX	0.87	4.12	1.20	0.43	-0.01 → 1.76	-3.23 → 4.98	-4.37 → -2.09	3.84 → 6.12
Flex	*θX	0.27	4.68	2.35	0.49	-0.74 → 1.28	-4.41 → 4.95	-5.65 → -3.17	3.71 → 6.19

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Zebris [mean HRA] minus Laser [mean HRA]); s<sub>c</sub> is the adjusted SD of the differences; s<sub>d</sub> is the SD of the differences without adjusted variation

\* Data sets exhibiting proportional bias &/or heteroscedasticity.

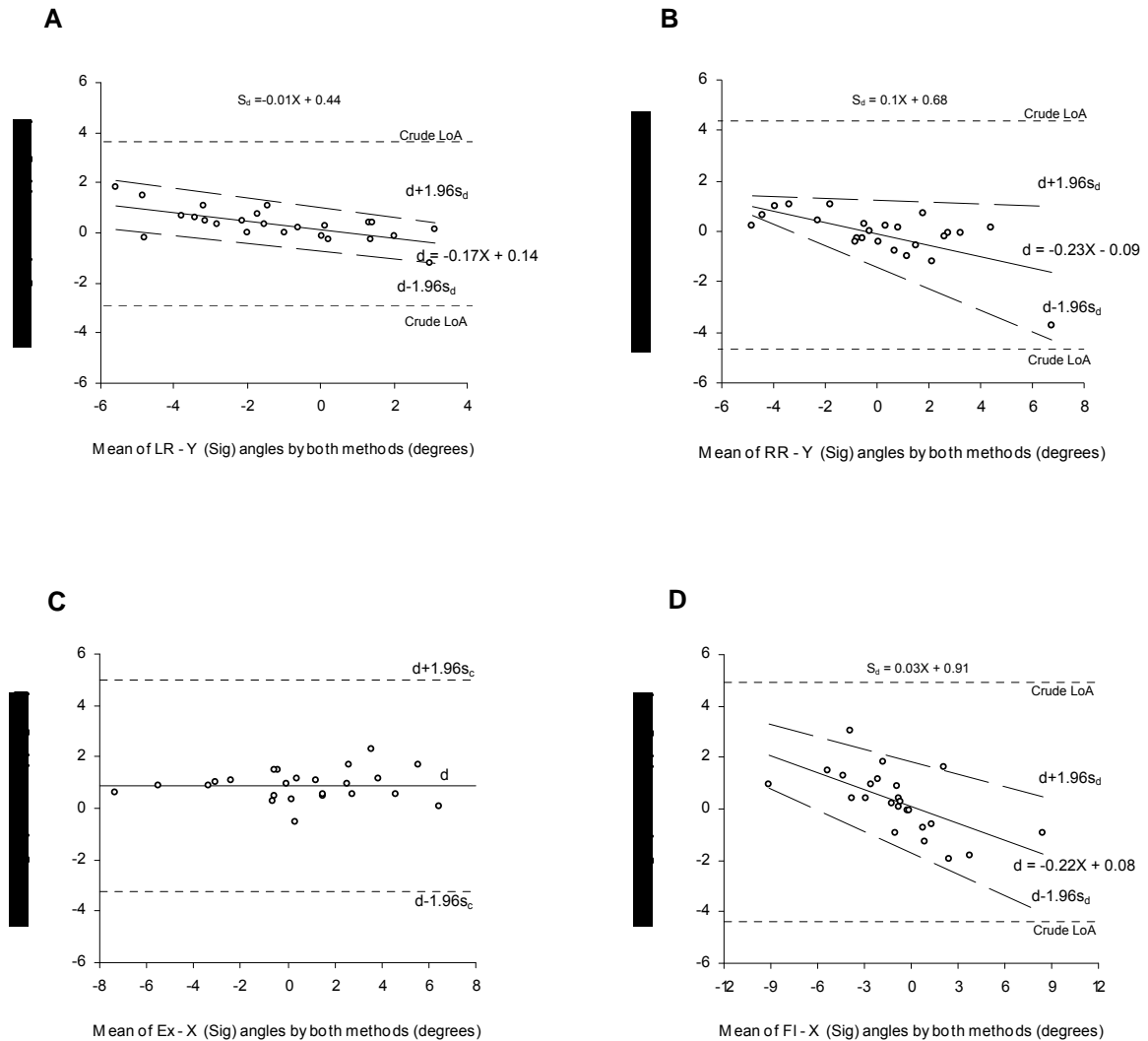
When the data were examined for relationships between the differences and the means, statistically significant *proportional biases* were seen with both examiners for all but the  $\text{Ext} \Rightarrow 0$  error trial. A statistically significant *heteroscedastic* relationship was also apparent for the L and RR  $\Rightarrow 0$  trials with examiner one and for the RR and Flex  $\Rightarrow 0$  trials with the second examiner (Figures 49 and 50). The adjusted bias and agreement estimates (regression analysis) for the affected data sets using are shown in Table 37.

### Examiner 1



**Figure 49.** Plots of the difference against mean (repeated measures) between Zebris and Laser methods by Examiner 1 for the signed data head repositioning accuracy (HRA). On each plot; systematic error or bias is indicated by the mean value line ( $d$ ); 'limits of agreement' (LoA) are given by the  $d \pm 1.96s_d$  lines. A - left rotation Y-axis; B - right rotation Y-axis; C - extension X-axis; D - flexion X-axis.

## Examiner 2



**Figure 50.** Plots of the difference against mean (repeated measures) between Zebris and Laser methods by Examiner 2 for the signed head repositioning accuracy (HRA). On each plot; systematic error or bias is indicated by the mean value line (d); 'limits of agreement' (LoA) are given by the  $d \pm 1.96S_d$  lines. A - left rotation Y-axis; B - right rotation Y-axis; C - extension X-axis; D - flexion X-axis.

All *proportional biases* were observed in the same negative direction and less than  $0.3^\circ$  the majority were  $0.2^\circ$  (Table 37). The relative effect of this bias on the overshoot/undershoot agreement between methods depends on the direction of primary movement for the trial. For all trials with data exhibiting a relationship, the negative *proportional bias* represents proportionally smaller laser pointer signed estimates for high negative values (undershoot for LR and Flex) and proportionally larger values for high positive scores (undershoot for

RR and Ext). As previously mentioned, the *fixed bias* estimates were not statistically affected by the *proportional bias*. Overall, when the results exhibiting only *proportional bias* were adjusted, there was a general small reduction in the LoA's (reduced *random error* estimates) together with a shift in the LoA's in the direction described above. When *proportional bias* and *heteroscedasticity* were present a comparison of the adjusted results with the unadjusted single session data revealed a similar shift in the LoA's and a relative narrowing of the LoA for larger negative values and widening for higher positive values.

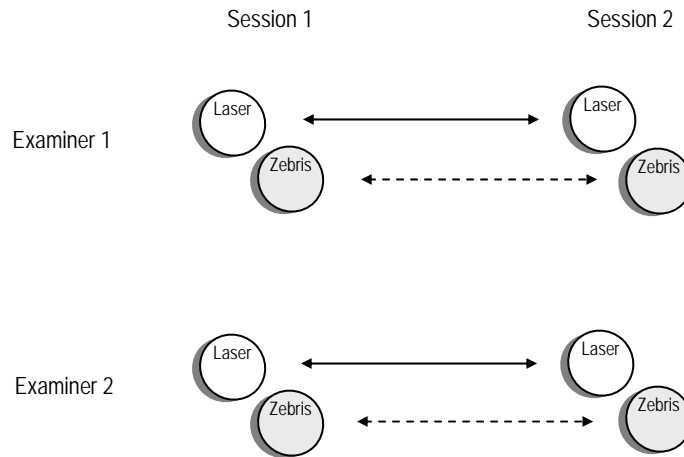
**Table 37.** Results of adjusted Bland and Altman analysis (regression approach) for method agreement (n=24) between the two different measurement methods (Zebris and Laser) with data sets exhibiting a relationship between the mean errors and differences.

Primary Motion	Examiner	Axis of Motion	Fixed Bias	95%CI for Fixed Bias	Prop bias	Highest negative mean HRA value		Highest positive mean HRA value	
						$d \pm 1.96s_d$	95% limits of agreement	$d \pm 1.96s_d$	95% limits of agreement
LR	1	$\theta Y$	0.1	-0.16 → 0.38	-0.2	$1.1 \pm 0.7$	0.4 → 1.8	$-0.6 \pm 1.5$	-2.1 → 0.9
	2	$\theta Y$	0.1	-0.90 → 0.37	-0.2	$1.1 \pm 1.0$	0.1 → 2.1	$-0.4 \pm 0.8$	-1.2 → 0.4
RR	1	$\theta Y$	-0.1	-0.44 → 0.18	-0.1	$1.0 \pm 0.2$	0.8 → 1.2	$-1.1 \pm 2.1$	-3.2 → 1.0
	2	$\theta Y$	-0.1	-0.41 → 0.23	-0.2	$1.0 \pm 0.4$	0.6 → 1.4	$-1.6 \pm 2.7$	-4.3 → 1.1
Flex	1	$\theta X$	-0.1	-0.50 → 0.26	-0.3	$1.5 \pm 1.9$	-0.4 → 3.4	$-2.2 \pm 1.6$	-3.8 → -0.6
	2	$\theta X$	0.1	-0.35 → 0.50	-0.2	$2.1 \pm 1.2$	0.9 → 3.3	$-1.8 \pm 2.3$	-4.1 → 0.5

*Abbreviations:* left and right rotation (LR, RR); Prop - proportional  
d is the mean of differences (Zebris - Laser means) ; a +ve value indicates Zebris > Laser.

## 2) One-week Test-Retest Precision for HRA-SSA: Laser and Zebris method

A visual examination of the distribution plots followed by a statistical evaluation of the relationship between the differences and the mean values (Pearson's *r* and absolute residual analysis), indicated there was no relationship between these variables with all data sets (absolute and signed). As such, the results are presented using the standard analysis approach. The test-retest precision was investigated for both methods of measurement and with each of the two examiners (Figure 51). Therefore, four sets of results are presented.



**Figure 51.** Overview of study design and data handling for analysis of test-retest precision

**One-week test-retest precision (laser and Zebris): absolute values.** The results of the Bland and Altman analysis for each examiner are shown in Tables 38 and 49 with the laser pointer method and in Tables 40 and 41 using the Zebris system. Estimates of *fixed bias* indicated a small but relatively consistent directional aspect (i.e., +ve difference indicates test > retest) with slightly larger measurements at the first (test) measurement session. For the majority of measurements (31 of 32), the estimates of *fixed bias* (d) were less than 1.0°. Overall, the results indicate that there was no statistically significant *fixed bias* (d) between the testing sessions with either method for both examiners. For both examiners with the laser pointer method, estimates of test-retest *random error* ( $\pm 1.96s_d$ ) around the axis of primary of motion for the L and RR  $\Rightarrow$  0 trials ( $\theta Y$ ) ranged from 3.7 to 4.8 degrees and from 4.2 to 4.5 degrees for the axis of primary of motion with the Flex and Ext  $\Rightarrow$  0 trials ( $\theta X$ ). With the Zebris system, estimates of *random error* ( $\pm 1.96s_d$ ) around the axis of primary of motion for the L and RR  $\Rightarrow$  0 trials ( $\theta Y$ ) ranged from 2.9 to 4.7 degrees and from 3.7 to 4.5 degrees for the axis of primary of motion with the Flex and Ext  $\Rightarrow$  0 trials ( $\theta X$ ). There was a tendency for the *random error* estimates to be slightly higher with examiner one using both measurement methods. For both examiners with both methods, the *random errors* for the Flex and Ext  $\Rightarrow$  0 trials, were generally relatively larger around the axis of primary of motion ( $\theta X$ ) compared with the perpendicular axis or rotation ( $\theta Y$ ).

**Table 38.** Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 1 using the laser pointer method.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.37	4.80	0.50	-1.40 → 0.66	-5.16 → 4.43	-6.84 → -3.49	2.75 → 6.10
	$\theta Y^\dagger$	-0.31	4.82	0.50	-1.35 → 0.73	-5.13 → 4.51	-6.81 → -3.45	2.83 → 6.19
RR	$\theta X$	0.44	3.45	0.36	-0.31 → 1.18	-3.02 → 3.89	-4.23 → -1.81	2.69 → 5.10
	$\theta Y^\dagger$	0.40	3.94	0.41	-0.45 → 1.25	-3.54 → 4.35	-4.92 → -2.16	2.97 → 5.72
Ext	$\theta X^\dagger$	0.02	4.41	0.46	-0.93 → 0.97	-4.40 → 4.44	-5.94 → -2.86	2.90 → 5.98
	$\theta Y$	-0.37	3.19	0.33	-1.05 → 0.32	-3.56 → 2.83	-4.68 → -2.45	1.72 → 3.95
Flex	$\theta X^\dagger$	0.12	4.51	0.47	-0.85 → 1.10	-4.39 → 4.63	-5.96 → -2.81	3.06 → 6.21
	$\theta Y$	0.41	4.04	0.42	-0.46 → 1.28	-3.62 → 4.45	-5.04 → -2.22	3.04 → 5.86

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement

d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 39.** Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 2 using the laser pointer method.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	0.66	3.55	0.37	-0.10 → 1.42	-2.89 → 4.21	-4.12 → -1.65	2.97 → 5.44
	$\theta Y^\dagger$	0.54	3.06	0.32	-0.12 → 1.21	-2.52 → 3.61	-3.60 → -1.45	2.54 → 4.69
RR	$\theta X$	0.31	3.21	0.33	-0.39 → 1.00	-2.91 → 3.52	-4.03 → -1.78	2.40 → 4.64
	$\theta Y^\dagger$	0.00	4.53	0.47	-0.84 → 0.84	-4.53 → 4.53	-6.11 → -2.95	2.95 → 6.12
Ext	$\theta X^\dagger$	-0.13	4.43	0.46	-1.08 → 0.83	-4.55 → 4.30	-6.10 → -3.01	2.76 → 5.85
	$\theta Y$	0.45	3.04	0.32	-0.21 → 1.10	-2.60 → 3.50	-3.66 → -1.54	2.43 → 4.56
Flex	$\theta X^\dagger$	0.89	4.16	0.43	-0.01 → 1.79	-3.27 → 5.05	-4.72 → -1.81	3.60 → 6.50
	$\theta Y$	0.77	2.74	0.29	0.18 → 1.36	-1.96 → 3.52	-2.94 → -1.02	2.56 → 4.48

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement

d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 40.** Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 1 using the Zebris system.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.33	3.31	0.34	-1.05 → 0.38	-3.64 → 2.98	-4.80 → -2.49	1.82 → 4.13
	$\theta Y^\dagger$	0.01	4.21	0.44	-0.90 → 0.92	-4.20 → 4.22	-5.68 → -2.73	2.75 → 5.69
RR	$\theta X$	0.07	3.33	0.35	-0.65 → 0.79	-3.26 → 3.42	-4.42 → -2.10	2.24 → 4.56
	$\theta Y^\dagger$	0.44	3.19	0.33	-0.25 → 1.12	-2.75 → 3.62	-3.87 → -1.64	2.51 → 4.74
Ext	$\theta X^\dagger$	0.52	4.47	0.47	-0.44 → 1.49	-3.95 → 4.99	-5.50 → -2.39	3.43 → 6.55
	$\theta Y$	-0.44	2.55	0.26	-0.99 → 0.10	-2.98 → 2.10	-3.87 → -2.10	1.21 → 2.98
Flex	$\theta X^\dagger$	0.06	4.14	0.43	-0.84 → 0.95	-4.08 → 4.19	-5.52 → -2.64	2.75 → 5.64
	$\theta Y$	0.55	2.94	0.31	-0.08 → 1.18	-2.39 → 3.48	-3.41 → -1.36	2.46 → 4.50

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 41.** Results of Bland and Altman analysis of test-retest precision for absolute (unsigned) head repositioning accuracy (degrees) obtained by Examiner 2 using the Zebris system.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.38	3.25	0.34	-1.08 → 0.32	-3.64 → 2.87	-4.77 → -2.50	1.73 → 4.00
	$\theta Y^\dagger$	0.47	2.86	0.30	-0.15 → 1.09	-2.40 → 3.34	-3.40 → -1.40	2.34 → 4.34
RR	$\theta X$	0.20	2.92	0.30	-0.43 → 0.82	-2.72 → 3.11	-3.73 → -1.70	2.09 → 4.13
	$\theta Y^\dagger$	0.15	4.72	0.49	-0.86 → 1.17	-4.57 → 4.88	-6.22 → -2.92	3.23 → 6.53
Ext	$\theta X^\dagger$	-0.07	3.65	0.38	-0.86 → 0.71	-3.71 → 3.57	-4.98 → -2.44	2.30 → 4.84
	$\theta Y$	0.51	2.08	0.22	0.06 → 0.96	-1.57 → 2.58	-2.29 → -0.84	1.86 → 3.31
Flex	$\theta X^\dagger$	1.07	4.45	0.46	0.11 → 2.03	-3.38 → 5.53	-4.94 → -1.83	3.97 → 7.08
	$\theta Y$	0.71	2.10	0.22	0.26 → 1.16	-1.39 → 2.81	-2.12 → -0.66	2.08 → 3.54

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.



**One-week test-retest precision (laser and Zebris): signed values.** The results of the Bland and Altman analysis for each examiner are shown in Tables 42 and 43 with the laser pointer method and in Tables 44 and 45 using the Zebris system.

**Table 42.** Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 1 using the laser pointer method.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	1.28	3.19	0.65	-0.07 → 2.62	-4.97 → 7.52	-7.15 → -2.79	5.34 → 9.70
RR	θY	-0.17	3.32	0.68	-1.57 → 1.23	-6.68 → 6.33	-8.95 → -4.41	4.06 → 8.60
Ext	θX	-0.14	2.97	0.61	-1.40 → 1.12	-5.97 → 5.69	-8.00 → -3.93	3.66 → 7.73
Flex	θX	-0.80	2.47	0.51	-1.84 → 0.25	-5.65 → 4.05	-7.34 → -3.95	2.36 → 5.74

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences.

**Table 43.** Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 2 using the laser pointer method.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	0.00	4.51	0.47	-0.97 → 0.97	-4.51 → 4.51	-6.09 → -2.94	2.93 → 6.08
RR	θY	0.68	5.63	0.59	-0.54 → 1.89	-4.95 → 6.30	-4.30 → -1.73	3.06 → 5.63
Ext	θX	-0.88	5.90	0.61	-2.15 → 0.39	-6.78 → 5.02	-8.84 → -4.72	2.96 → 7.08
Flex	θX	-1.05	4.96	0.52	-2.12 → 0.02	-6.01 → 3.91	-7.75 → -4.28	2.18 → 5.65

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences.

**Table 44.** Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 1 using the Zebris system.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	1.12	5.39	0.56	-0.04 → 2.27	-4.27 → 6.50	-6.15 → -2.39	4.62 → 8.38
RR	θY	0.12	5.15	0.54	-0.99 → 1.23	-5.03 → 5.27	-6.82 → -3.32	3.48 → 7.07
Ext	θX	0.09	5.03	0.59	-1.00 → 1.17	-4.95 → 5.13	-6.71 → -3.19	3.37 → 6.89
Flex	θX	-0.75	4.74	0.49	-1.77 → 0.28	-5.49 → 4.00	-7.15 → -3.84	2.35 → 5.66

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences.

**Table 45.** Results of Bland and Altman analysis of test-retest precision for signed head repositioning accuracy (degrees) obtained by Examiner 2 using the Zebris system.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	0.08	4.23	0.44	-0.83 → 0.99	-4.15 → 4.31	-5.63 → -2.67	2.83 → 5.78
RR	θY	0.82	5.74	0.60	-0.42 → 2.06	-4.93 → 6.56	-6.93 → -2.92	4.56 → 8.57
Ext	θX	-1.00	4.74	0.49	-2.02 → 0.02	-5.74 → 3.74	-7.40 → -4.09	2.09 → 5.40
Flex	θX	-1.39	5.61	0.58	-2.60 → -0.19	-7.00 → 4.21	-8.96 → -5.04	2.25 → 6.17

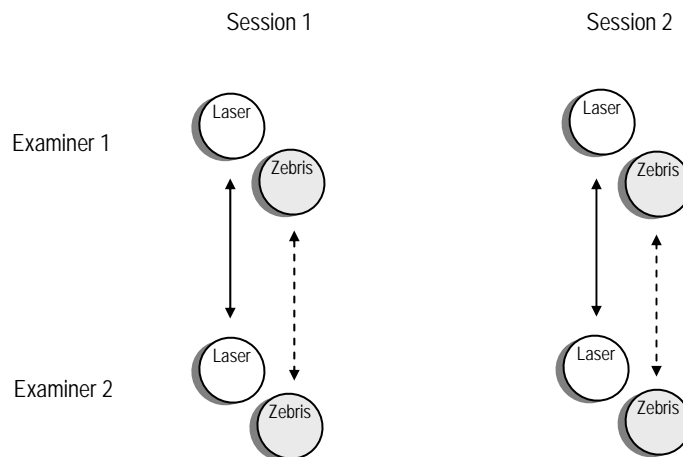
*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Test - Retest means); s<sub>d</sub> is the SD of the differences.

An examination for *fixed bias* (d) showed that for LR and Ext ⇒ 0 trials, the measurements with both methods and examiners tended toward a relative undershoot at the retest session compared with the first session. For the RR and Flex ⇒ 0 trials, the measurements tended toward a relative overshoot at the retest session. However, overall the mean differences (d) between testing sessions were less than 1.4° for all error components and the results indicate that there was no consistent statistically significant *fixed bias* (d) between the testing sessions for both examiners with either method. For the laser pointer method, estimates of

test-retest *random error* ( $\pm 1.96s_d$ ) were consistently smaller for examiner one with L and RR  $\Rightarrow 0$  ranging from 3.2 to 3.3 degrees compared with 4.5 to 5.6 with the examiner two. For Flex and Ext  $\Rightarrow 0$  trials, estimates of *random error* for examiner one ranged from 2.5 to 3.0 degrees compared with 5.0 to 5.9 with the examiner two. This pattern was not observed with the Zebris system. For the Zebris method with both examiners, the estimates of test-retest *random error* for the signed error components ranged from 4.2 to 5.7 degrees for the L & RR  $\Rightarrow 0$  trials and 4.7 to 5.6 degrees for the Flex and Ext  $\Rightarrow 0$  trials

### 3) Immediate Inter-Examiner Precision for HRA-SSA: Laser and Zebris Method

A visual examination of the distribution plots followed by a statistical evaluation the relationship between the differences and the mean values (Pearson's  $r$  and absolute residual analysis), indicated there was no relationship between these variables with all data sets (absolute and signed). As such, the results are presented using the standard analysis approach. The inter-examiner precision was investigated for both methods at both the test and retest sessions (Figure 52). Although there was no clear difference in the results between sessions, all four sets of results are presented.



**Figure 52.** Overview of study design and data handling for analysis

**Immediate Inter-Examiner Precision (laser and Zebris): absolute values.** The results of the Bland and Altman analysis for each session are shown in Tables 46 and 47 with the laser pointer method and in Tables 48 and 49 with the Zebris system. Overall, the results indicate that there was no consistent statistically significant *fixed bias* ( $d$ ) between the examiners with either method at both testing sessions. There was no consistent directional (i.e., +ve difference indicates examiner one > examiner two) bias to any differences observed and for the majority of measurements (31 of 32), the estimates of test-retest *fixed bias* ( $d$ ) were less than  $1.0^\circ$ . Using the results from both testing sessions, estimates of inter-examiner *random error* ( $\pm 1.96s_d$ ) with the laser pointer method around the axis of primary of motion ranged from 3.7 to 4.2 degrees for the L and RR  $\Rightarrow$  0 trials ( $\theta Y$ ) and from 4.5 to 5.5 degrees for the axis of primary of motion with the Flex and Ext  $\Rightarrow$  0 trials ( $\theta X$ ). With the Zebris system, estimates of *random error* ( $\pm 1.96s_d$ ) around the axis of primary of motion for the L and RR  $\Rightarrow$  0 trials ( $\theta Y$ ) ranged from 3.2 to 4.1 degrees and from 3.7 to 5.1 degrees for the axis of primary of motion with the Flex and Ext  $\Rightarrow$  0 trials ( $\theta X$ ). For both methods of measurement at each session, there was a difference in the inter-examiner *random error* between the axis of primary of motion and the perpendicular axis for the Flex and Ext  $\Rightarrow$  0 trials compared with the L and RR  $\Rightarrow$  0 trials. For the Flex and Ext  $\Rightarrow$  0 trials, the inter-examiner *random errors* were generally larger in the plane of primary motion.

**Table 46.** Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session I using the laser pointer method.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.47	3.02	0.31	-1.13 → 0.18	-3.50 → 2.55	-4.55 → -2.44	1.49 → 3.60
	$\theta Y^\dagger$	0.20	3.70	0.39	-0.59 → 1.00	-3.50 → 3.90	-4.79 → -2.20	2.61 → 5.19
RR	$\theta X$	0.24	3.80	0.40	-0.58 → 1.06	-3.57 → 4.05	-4.90 → -2.24	2.72 → 5.38
	$\theta Y^\dagger$	0.67	3.68	0.38	-0.13 → 1.46	-3.02 → 4.35	-4.30 → -1.73	3.06 → 5.63
Ext	$\theta X^\dagger$	-0.35	5.45	0.57	-1.52 → 0.83	-5.79 → 5.09	-7.69 → -3.89	3.19 → 6.99
	$\theta Y$	-0.30	2.74	0.28	-0.89 → 0.29	-3.03 → 2.44	-3.99 → -2.08	1.48 → 3.39
Flex	$\theta X^\dagger$	-0.79	4.49	0.47	-1.76 → 0.18	-5.29 → 3.71	-6.86 → -3.72	2.14 → 5.28
	$\theta Y$	0.08	3.98	0.41	-0.78 → 0.94	-3.90 → 4.06	-5.29 → -2.51	2.67 → 5.45

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
 $d$  is the mean of differences (Examiner one – Examiner two means);  $s_d$  is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 47.** Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session II using the laser pointer method.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	0.55	3.65	0.38	-0.23 → 1.34	-3.09 → 4.20	-4.37 → -1.82	2.93 → 5.47
	$\theta Y^\dagger$	1.06	3.86	0.40	0.23 → 1.89	-2.81 → 4.93	-4.16 → -1.46	3.58 → 6.27
RR	$\theta X$	0.11	3.47	0.36	-0.64 → 0.85	-3.36 → 3.57	-4.56 → -2.15	2.36 → 4.78
	$\theta Y^\dagger$	0.26	4.17	0.43	-0.64 → 1.16	-3.91 → 4.44	-5.37 → -2.45	2.98 → 5.90
Ext	$\theta X^\dagger$	-0.49	4.96	0.52	-1.56 → 0.58	-5.46 → 4.48	-7.19 → -3.73	2.74 → 6.21
	$\theta Y$	0.52	3.29	0.34	-0.19 → 1.23	-2.78 → 3.81	-3.93 → -1.63	2.66 → 4.96
Flex	$\theta X^\dagger$	-0.02	4.45	0.46	-0.98 → 0.94	-4.48 → 4.43	-6.03 → -2.92	2.88 → 5.99
	$\theta Y$	0.44	2.37	0.25	-0.07 → 0.95	-1.92 → 2.80	-2.75 → -1.10	1.98 → 3.63

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 48.** Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session I using the Zebri system.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.01	2.41	0.25	-0.53 → 0.51	-2.41 → 2.40	-3.26 → -1.57	1.56 → 3.24
	$\theta Y^\dagger$	0.24	3.18	0.33	-0.45 → 0.93	-2.94 → 3.42	-4.05 → -1.83	2.31 → 4.54
RR	$\theta X$	-0.04	3.60	0.38	-0.81 → 0.74	-3.65 → 3.58	-4.91 → -2.39	2.32 → 4.84
	$\theta Y^\dagger$	0.66	3.25	0.34	-0.04 → 1.37	-2.62 → 3.95	-3.77 → -1.48	2.80 → 5.10
Ext	$\theta X^\dagger$	-0.18	5.12	0.53	-1.29 → 0.92	-5.31 → 4.94	-7.10 → -3.52	3.15 → 6.73
	$\theta Y$	-0.33	2.27	0.24	-0.82 → 0.16	-2.59 → 1.93	-3.38 → -1.80	1.14 → 2.73
Flex	$\theta X^\dagger$	-0.92	4.49	0.47	-1.89 → 0.04	-5.40 → 3.55	-6.97 → -3.84	1.99 → 5.12
	$\theta Y$	0.20	3.02	0.31	-0.45 → 0.85	-2.82 → 3.21	-3.87 → -1.77	2.16 → 4.26

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

<sup>†</sup> Primary axis of rotation for the repositioning test.

**Table 49.** Results of Bland and Altman analysis of Inter-Examiner precision for absolute (unsigned) head repositioning accuracy (degrees) at session II using the Zebris system.

Trial	Axis of Rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	$\theta X$	-0.06	3.27	0.34	-0.76 → 0.65	-3.32 → 3.21	-4.46 → -2.18	2.07 → 4.34
	$\theta Y^\dagger$	0.70	3.82	0.40	-0.12 → 1.53	-3.12 → 4.53	-4.46 → -1.79	3.19 → 5.86
RR	$\theta X$	0.09	3.41	0.35	-0.64 → 0.82	-3.31 → 3.50	-4.50 → -2.13	2.31 → 4.69
	$\theta Y^\dagger$	0.38	4.12	0.43	-0.50 → 1.27	-3.74 → 4.51	-5.18 → -2.30	3.07 → 5.94
Ext	$\theta X^\dagger$	-0.78	4.37	0.45	-1.72 → 0.16	-5.14 → 3.58	-6.67 → -3.62	2.06 → 5.11
	$\theta Y$	0.62	2.33	0.24	0.12 → 1.13	-1.71 → 2.96	-2.53 → -0.90	2.14 → 3.77
Flex	$\theta X^\dagger$	0.09	3.74	0.39	-0.71 → 0.90	-3.66 → 3.85	-4.97 → -2.35	2.54 → 5.16
	$\theta Y$	0.36	1.76	0.18	-0.02 → 0.74	-1.40 → 2.12	-2.02 → -0.79	1.50 → 2.73

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement

d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

† Primary axis of rotation for the repositioning test.

**Immediate Inter-Examiner Precision (laser and Zebris): signed values.** The results of the Bland and Altman analysis for each session are shown in Tables 50 and 51 with the laser pointer method and in Tables 52 and 53 with the Zebris system. An examination for *fixed bias* (d) showed that for the Flex  $\Rightarrow$  0 trials, the measurements with both methods at both sessions, tended toward a relative overshoot with examiner two compared with the first examiner. No consistent directional bias pattern was seen between examiners with either method for the L, RR and Ext  $\Rightarrow$  0 trials when both sessions are considered. However, overall the mean differences (d) between examiners were less than 1.1° for all error components and the results indicate that there was no statistically significant *fixed bias* (d) between the examiners with either method at both sessions. Using data from both sessions, estimates of inter-examiner *random error* ( $\pm 1.96s_d$ ) for the laser pointer method ranged from 4.4 to 6.5 degrees for the L & RR  $\Rightarrow$  0 trials and 5.0 to 6.6 degrees for the Flex & Ext  $\Rightarrow$  0 trials. (Table 50 and Table 51). For the Zebris system, the estimates of inter-examiner *random error* ranged from 4.3 to 5.9 degrees for the L & RR  $\Rightarrow$  0 trials and 5.0 to 6.1 degrees for the Flex & Ext  $\Rightarrow$  0 trials (Tables 52 and 53).

**Table 50.** Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 1 using the laser pointer method.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	0.47	5.29	0.55	-0.67 → 1.61	-4.82 → 5.75	-6.66 → -2.97	3.91 → 7.60
RR	θY	-0.57	6.45	0.67	-1.96 → 0.82	-7.02 → 5.88	-9.27 → -4.77	3.62 → 8.13
Ext	θX	1.11	6.29	0.66	-0.25 → 2.47	-5.18 → 7.40	-7.38 → -2.99	5.21 → 9.60
Flex	θX	0.81	5.00	0.52	-0.26 → 1.89	-4.18 → 5.81	-5.92 → -2.44	4.07 → 7.55

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

**Table 51.** Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 2 using the laser pointer method.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	-0.81	6.41	0.67	-2.19 → 0.57	-7.21 → 5.59	-9.44 → -4.97	3.36 → 7.83
RR	θY	0.28	4.43	0.46	-0.68 → 1.23	-4.16 → 4.71	-5.71 → -2.61	3.16 → 6.26
Ext	θX	0.37	6.55	0.68	-1.04 → 1.78	-6.18 → 6.92	-8.46 → -3.89	4.63 → 9.20
Flex	θX	0.56	5.53	0.58	-0.63 → 1.76	-4.97 → 6.10	-6.90 → -3.04	4.17 → 8.03

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

**Table 52.** Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 1 using the Zebris system.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	0.34	4.39	0.46	-0.61 → 1.29	-4.06 → 4.74	-5.60 → -2.53	3.20 → 6.27
RR	θY	-0.50	5.45	0.57	-1.67 → 0.68	-5.94 → 4.95	-7.84 → -4.04	3.05 → 6.85
Ext	θX	1.00	6.13	0.64	-0.32 → 2.32	-5.14 → 7.13	-7.28 → -3.00	4.99 → 9.28
Flex	θX	0.72	5.06	0.53	-0.38 → 1.81	-4.35 → 5.78	-6.12 → -2.58	4.01 → 7.55

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

**Table 53.** Results of Bland and Altman analysis of Inter-Examiner precision for signed head repositioning accuracy (degrees) at session 2 using the Zebris system.

Primary Motion	Axis of rotation	d	1.96s <sub>d</sub>	SE(d)	95% CI for d	Limits of agreement (LoA)		
						95% limits of agreement	95% CI for lower LoA	95% CI for upper LoA
LR	θY	-0.70	5.86	0.61	-1.96 → 0.56	-6.56 → 5.16	-8.61 → -4.51	3.12 → 7.12
RR	θY	0.20	4.29	0.45	-0.73 → 1.12	-4.09 → 4.49	-5.59 → -2.60	2.99 → 5.98
Ext	θX	-0.10	5.49	0.57	-1.27 → 1.09	-5.58 → 5.40	-7.50 → -3.66	3.48 → 7.31
Flex	θX	0.07	5.04	0.52	-1.02 → 1.15	-4.97 → 5.10	-6.73 → -3.21	3.35 → 6.86

*Abbreviations:* left and right rotation (LR, RR); extension and flexion (Ext, Flex); LoA – Limits of Agreement  
d is the mean of differences (Examiner one – Examiner two means); s<sub>d</sub> is the SD of the differences

### 5.3.3 Summary of Kinaesthetic Sensibility Primary Results.

#### 1) Concurrent Method agreement for HRA-SSA; Laser and Zebris method

- **Absolute HRA**

- 6 of the 16 repeated data sets exhibited *heteroscedasticity* &/or *proportional bias*. The majority of these were for error components in the plane of movement and seen with examiner two.
- No statistically significant *fixed bias* (d) between the two methods of measurement using the standard and regression adjusted results.
- For error components in the plane of movement, crude estimates of *random error* without repeated measurements ( $\pm 1.96s_d$ ) were slightly larger for examiner two. For both examiners *random error* estimates ranged from 1.2 to 2.7 degrees for the θY L and RR ⇒ 0 trials. The θX *random errors* for the Flex & Ext trials ranged and 1.7 to 2.8 degrees.
- The regression adjusted results exhibited smaller *random error* estimates than those without repeated measures.
- The incorporation of test-retest variance did result in consistently larger estimates of *random error* ( $\pm 1.96s_c$ ) but no change in estimates of *fixed bias*. For both examiners, *random error* estimates ranged from 2.5 to 3.6 degrees for the θY L and RR ⇒ 0 trials. The θX *random errors* for the Flex and Ext trials ranged and 3.4 to 3.8 degrees.



- **Signed HRA (overshoot/undershoot)**

- 6 of the 8 repeated data sets exhibited *heteroscedasticity &/or proportional bias*.
- No statistically significant *fixed bias* (d) between the two methods of measurement using the standard and regression adjusted results.
- For both examiners *random error* estimates without repeated measurements ( $\pm 1.96s_d$ ), ranged from 1.2 to 1.9 degrees for the L and RR  $\Rightarrow$  0 trials. The *random errors* for the Flex and Ext  $\Rightarrow$  0 trials ranged and 1.2 to 2.4 degrees.
- The regression adjusted results exhibited smaller *random error* estimates than those without repeated measures.
- The incorporation of test-retest variance did result in consistently larger estimates of *random error* ( $\pm 1.96s_c$ ) but no change in estimates of *fixed bias*. For both examiners, *random error* estimates ranged from 3.3 to 4.6 degrees for the  $\theta Y$  L and RR  $\Rightarrow$  0 trials. The  $\theta X$  *random errors* for the Flex and Ext trials ranged and 4.1 to 4.7 degrees.

## 2) One-week Test-Retest Precision for HRA-SSA; Laser and Zebris method

- **Absolute HRA**

- No statistically significant *fixed bias* (d) between the two testing sessions for either method.
- For the laser pointer method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 3.1 to 4.8 degrees for the  $\theta Y$  L and RR  $\Rightarrow$  0 trials. The  $\theta X$  *random errors* for the Flex & Ext trials ranged and 4.2 to 4.5 degrees.
- For the Zebris method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 2.9 to 4.2 degrees for the  $\theta Y$  L and RR  $\Rightarrow$  0 trials. The  $\theta X$  *random errors* for the Flex and Ext trials ranged and 3.7 to 4.5 degrees.

- **Signed HRA (overshoot/undershoot)**
  - No statistically significant *fixed bias* ( $d$ ) between the two testing sessions for either method.
  - Estimates of test-retest *random error* without repeated measurements ( $\pm 1.96s_d$ ) were larger for examiner 2 using the laser pointer method.
  - For the laser pointer method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 3.2 to 5.6 degrees for the L & RR  $\Rightarrow$  0 trials. The *random errors* for the Flex and Ext  $\Rightarrow$  0 trials ranged and 2.5 to 5.9 degrees.
  - For the Zebris method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 4.2 to 5.7 degrees for the L and RR  $\Rightarrow$  0 trials. The *random errors* for the Flex and Ext  $\Rightarrow$  0 trials ranged and 4.7 to 5.6 degrees.

### 3) Immediate Inter-Examiner Precision for HRA-SSA: Laser and Zebris Method

- **Absolute HRA**
  - No statistically significant *fixed bias* ( $d$ ) between the two testing sessions for either method.
  - For the laser pointer method with both sessions, the estimates of inter-examiner *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 3.7 to 4.2 degrees for the  $\theta Y$  L and RR  $\Rightarrow$  0 trials. The  $\theta X$  *random errors* for the Flex and Ext trials ranged and 4.5 to 5.5 degrees.
  - For the Zebris method with both sessions, the estimates of inter-examiner *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 3.2 to 4.1 degrees for the  $\theta Y$  L and RR  $\Rightarrow$  0 trials. The  $\theta X$  *random errors* for the Flex and Ext trials ranged and 3.7 to 5.1 degrees.

- **Signed HRA (overshoot/undershoot)**
  - No statistically significant *fixed bias* (d) between the two testing sessions for either method.
  - For the laser pointer method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 4.4 to 6.5 degrees for the L and RR  $\Rightarrow$  0 trials. The *random errors* for the Flex and Ext  $\Rightarrow$  0 trials ranged and 5.0 to 6.6 degrees.
  - For the Zebris method with both examiners, the estimates of test-retest *random error* for error components in the plane of movement (around the primary axis of rotation) ranged from 4.3 to 5.9 degrees for the L and RR  $\Rightarrow$  0 trials. The *random errors* for the Flex and Ext  $\Rightarrow$  0 trials ranged and 5.0 to 6.1 degrees.

## 5.4 DISCUSSION

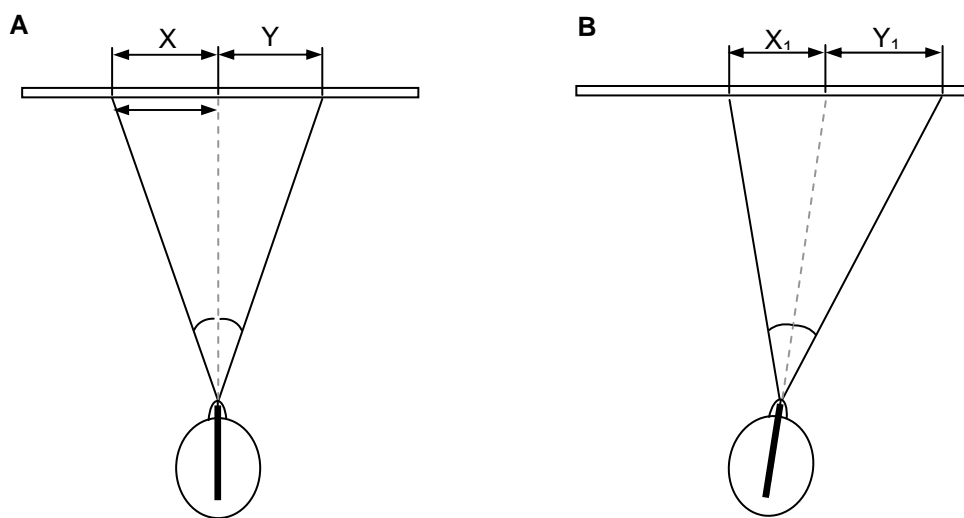
### Concurrent Method Agreement for HRA-SSA; Laser and Zebris method.

The purpose of this part of the study was to see if the methods of measurements using the laser pointer and Zebris system agree well enough for one method to replace another or perhaps for the two methods to be used interchangeably. The overall results from this study suggest that the two methods of measurement suggest that the methods may not agree sufficiently well for the methods to be used interchangeably. In discussing this further, a closer examination of the error and bias types and is again needed.

If the absolute data is considered, the results show that for nearly half of the method agreement HRA-SSA comparisons, *proportional bias* and/or *heteroscedasticity* was present. In contrast to the previous study (Chapter 4), all of these relationships were exhibited with the error components in the plane of movement (primary axis of rotation). There was also no clear relationship of outlying data to these errors/biases (Appendix IV). In addition, these errors/biases were more commonly seen with method agreement comparisons for examiner two, the less experienced investigator. When the distribution plots are observed (Figures 47 and 48), the *heteroscedastic* and *proportional* relationships are similar in nature and direction for each of the data sets. The *heteroscedastic* relationship signifies an increasing *random error* with progressively larger error measurements.

The negative *proportional bias* (multiply bias;  $Y = -bX$ ) indicates as the repositioning error increases, proportionally larger HRA values are seen with the laser pointer compared with the Zebris system. The similar nature of this *proportional bias* between the data sets suggests that this bias may arise from a consistent systematic source. Although it is not possible to accurately attribute the *bias* to either the laser pointer method or the Zebris system, it is most likely related to factors associated with the laser pointer technique. As previously highlighted, whilst the laser pointer method utilizes a relatively simple equipment design which is inexpensive, easy to execute and may permit a degree of discriminant classification of certain cervicalgic subgroups (Table 1 – Chapter 1), the method of measurement and in particular, its subjective and non-remote nature may involve a degree of experimenter bias and inaccuracy. However, the possible nature of the systematic bias seen in the current study, could relate to geometric factors when using a flat

target to record the repositioning errors when the head moves in an arc (Humphreys and Irgens, 2002). These factors are particularly relevant when the initial SSA position derived by the patient/subject is considered. Throughout the studies for this thesis, the initial SSA position was often not straight ahead with respect to the shoulders. This does not affect the Zebris measurements as the system is calibrated to zero at the SSA and subsequent measurements are angles in space around this target. However, for the laser pointer method, a head SSA position which is not truly straight ahead with respect to the shoulders will result in geometric inaccuracies (Figure 53).



**Figure 53.** Superior view of horizontal ( $\theta Y$ ) head repositioning using the laser pointer. A) distances X & Y are the same when the initial head position (SSA) is perpendicular to target. B) Same angle of repositioning as with A) but  $X_1$  &  $Y_1$  now altered from X & Y, giving erroneous angle from the laser pointer method.

The diagrams in Figure 53 show an overhead view of the laser pointer method with the horizontal plane repositioning trials (L and RR  $\Rightarrow$  0). Figure 53A shows the subject in a true straight ahead position with respect to the shoulders/target and therefore the laser pointer hitting the target at a  $90^\circ$  angle from a distance of 90 cm. In this position, distance X equals distance Y and the angles measured by both the Zebris and laser pointer method should be congruent (aside from other sources of variance). If the initial SSA position is rotated with respect to the trunk (Figure 53B), the same angle as in Figure 53A will be measured by the Zebris system. However the true laser to target distance will be longer than 90 cm and  $X_1$  and  $Y_1$  will be inaccurate. This will result in inaccurate angles with the laser

pointer method by an amount proportional to the head deviation from SSA, and the size of the true error. The same geometric argument also applies to the vertical repositioning errors. An argument against this geometric issue is the observation that the *proportional bias* was not seen in the error component perpendicular to the plane of motion for the trial. As the investigator did not control for this, it is reasonable to assume that all error components would be affected. As highlighted, the Zebris device is not dependent on these geometric principles and as such could be viewed as a criterion instrument. A study comparing the Zebris system to a digital inclinometer have shown ‘good’ agreement (Dvir and Prushansky, 2000). However, the analysis of agreement did not include the most appropriate methods. In addition, this agreement conclusion was tested at ranges of motion compatible and set in context with actual cervical motion rather than those smaller angles seen with HRA. Only one study using the laser pointer method addressed the trigonometric issue arising from the initial SSA position of the patient/subject (Palmgren et al., 2006). With their testing method, Palmgren et al., (2006) aimed to make sure the beam from laser pointer was at right angles to the target (frontal plane) and also parallel to floor. However, with this method of measurement, the target head position would no longer be the subjective neutral (or straight ahead) position of the patient. To assess the effect of these geometric issues on the method agreement, the study could be repeated but with the laser pointer initial position being at right angles to the target both in the frontal and sagittal plane.

The use of concurrent measurements to assess method agreement was adopted for this study so to allow a comparison of methods without including any test-retest biological variability (subject variability in performing the HRA test). Any variability seen could therefore only be attributed to the following sources:

- Instrument error/variability (*i.e., the variability of the measurement device*).
- Rater error/variability (*i.e., the variability of the researcher/s or clinician/s administering the measurement device*).

Using this study design, the method agreement *random error* was found to be relatively small (approx  $\pm 1^\circ - 3^\circ$  without regression adjustment). With regression adjustment, these estimates were smaller. The acceptability of this level of agreement is dependent on the intended use of the measurement method and the size of values under investigation. In

essence the unadjusted results suggest that if a subject was tested with both methods, we could expect the HRA measurements to vary randomly by up to approximately  $\pm 3^\circ$ . If the relative size of errors that have been found in normal subjects and patients are considered (Tables 1-3), the *random error* coupled with other error/biases seen with this study, may be unacceptably large for the methods to be used interchangeably.

For the signed data, the results show that for nearly all of the method agreement HRA-SSA comparisons, *proportional bias* and/or *heteroscedasticity* was present. The reasons for this are likely the same as for the absolute data. The unadjusted results show that if an individual subject was tested with with both methods, we can expect the over/undershoot estimates to vary randomly by up to  $\pm 2.5^\circ$ . With the relatively small over/undershoot score seen in this study (Table 31) this means that using the two methods a subjects over/undershoot score may randomly vary from a relative overshoot to undershoot position or vice versa. As with the absolute measurements, these errors may be unacceptably large for the methods to be used interchangeably.

### **One-week Test-Retest Precision for HRA-SSA: Laser and Zebris method**

The purpose of this part of the study was to investigate the variability between measurements using both methods taken one week apart. If the absolute and signed data are considered, the results for both examiners with each method, revealed that the variability between measurements consisted of *random error* only. The *random errors* were similar between examiners and methods. For the absolute errors, in essence the results suggest that if a subject was tested on two occasions, one week apart using either method, we could expect the HRA measurements to vary randomly by up to approximately  $\pm 5^\circ$  for L and RR  $\Rightarrow 0$  in the plane of movement ( $\theta Y$ -axis) and approximately  $\pm 4.5^\circ$   $\theta X$ -axis components for the Flex and Ext trials. As highlighted earlier, the acceptability of this level of agreement depends on the intended use of the tests. If the HRA results are considered with respect to the  $4.5^\circ$  radial threshold originally described by Revel et al., (1991), the level of agreement shown in this study would not be acceptable when looking to differentiate neck pain from healthy subjects or monitoring change in patients over the course of a treatment plan.

For the over/undershoot characteristics, the results suggest that if a subject was tested on two occasions, one week apart using either method, we could expect the HRA measurements to vary randomly by up to approximately  $\pm 5.5^\circ$  for both the L and RR  $\Rightarrow 0$  and Flex and Ext trials in the plane of movement. Again, with the relatively small over/undershoot score seen in this study (Table 31) this means that between the two testing sessions, the subjects over/undershoot score may randomly vary from a relative overshoot to undershoot position or vice versa.

As shown in Table 28, several other studies have reported on test-retest precision for HRA measurement to a NHP or SSA (Revel et al., 1991; Heikkila and Astrom, 1996; Kristjansson et al., 2001; Lee et al., 2006; Swait et al., 2007). As with most of the cervicocephalic sensibility studies, direct comparison of the results is hampered by differences in factors such as testing instruments and protocol and for the *reliability* studies, specifically differences in testing intervals and statistical analyses (Table 28). The earliest studies used hypothesis tests (statistically significant differences) Some studies have used the ICC as a measure of agreement (Kristjansson et al., 2001; Lee et al., 2006; Swait et al., 2007). As highlighted in Chapter 3, this statistical approach has several important limitations especially when used in isolation. If the current study data (around the primary axis of motion) is analysed with an *absolute* ICC (2,5) model, the values range from 0.18 to 0.84 with the majority (11 of 16) greater than 0.5. These are similar to those seen with the two of other studies but lower than Swait et al., (2007). Both Kristjansson et al., (2001) and Swait et al., (2007) used the same ICC model as for this study (model 2), Kristjansson et al., (2001) used a single measures version (ICC(2,1)) whereas Swait et al., 2007 used an average measures version (ICC(2,k)). In addition, it's not clear whether an *absolute* or *consistency* version was used by the investigators (Kristjansson et al., 2001; Swait et al., 2007). However, as only *random error* was present between the measurements, both versions are likely to give the same ICC value (Table 12 – Chapter 3). Lee et al., (2006) used a one-way effects model (ICC(1,3)). Used on the same data, these tend to yield less generous ICC values than the two-way effects models (Table 12 – Chapter 3) (Haas, 1991a; Haas, 1991b).

As discussed in Chapter 3, the ICC should ideally be used in conjunction with agreement measures such as the Bland and Altman method (Bland and Altman, 1986), that allow a quantification of the nature and size of *random error* and *systematic bias* present in the



study. Only one other study complemented their ICC analysis with the Bland and Altman analysis (Kristjansson et al., 2001). However, they not provide adequate analysis outcomes that would allow a comparison with the current study.

### **Immediate Inter-Examiner Precision for HRA-SSA: Laser and Zebris Method**

The purpose of this part of the study was to investigate the variability between measurements taken by two examiners using both methods. As with the test-retest study, for the absolute and signed data, the variability between both examiners using each method, consisted of *random error* only. The magnitude of these errors was similar at each session and with each method. If the absolute data are considered, the results suggest that if a subject was tested by one examiner and then immediately by another, we could expect the HRA measurements to vary randomly by approximately  $\pm 4.5^\circ$  for L and RR  $\Rightarrow$  0 in the plane of movement ( $\theta Y$ -axis) and approximately  $\pm 5.5^\circ$   $\theta X$ -axis components for the Flex and Ext trials. This level of error is similar to that seen with test-retest precision. Therefore, the introduction of another examiner does not increase the variability and as such the same subject would not need to be tested by the same examiner each time. The arguments for the acceptability of this level of error are as before, dependent on the intended use of the measurements.

If the signed data are considered, the results suggest that if a subject were tested by one and then another examiner, we could expect the HRA measurements to vary randomly by up to approximately  $\pm 6.5^\circ$  for both the L and RR  $\Rightarrow$  0 and Flex and Ext trials in the plane of movement. As with the test-retest data, the subjects over/undershoot results could vary randomly from a relative overshoot to undershoot or vice versa.

Only one other study has reported on inter-rater precision (*reliability*) for measurement to a NHP or SSA (Revel et al., 1991). However, this study used hypothesis tests (statistically significant differences) to measure agreement and therefore a comparison with the current study is not possible.

## 5.5 CONCLUSION

Overall, the results show that the Zebris and laser pointer methods may not agree sufficiently well for them to be used interchangeably. The decision on acceptability of this level of agreement is ultimately dependent on the intended use. Bench validity testing (e.g., against a digital inclinometer) may give useful data as to which should be used as a criterion device. The test-retest precision was comparable between both methods suggesting that from this perspective either could be used for measuring HRA-SSA. The inter-rater test-retest precision was comparable to the test-retest precision suggesting that suitably trained examiners could be interchanged when carrying out repeated measurements. The level of agreement particularly *random error* as indicated by the Bland and Altman analysis, should be taken into consideration when looking to discriminate subjects with HRA deficits from normal and when monitoring change in HRA.

## **Chapter 6**

---

# **HEAD REPOSITIONING ACCURACY IN PATIENTS WITH CHRONIC, TRAUMATIC AND INSIDIOUS ONSET NECK PAIN: A STUDY OF THREE REPOSITIONING TESTS.**

---

### **6.1 INTRODUCTION**

The earlier study in this thesis comparing HRA-SSA in patients with insidious onset chronic neck pain to healthy control subjects failed to demonstrate any clinically meaningful difference between the groups (Chapter 2). These results were consistent with Teng et al., (2007) in patients with previous (subclinical/ chronic recurrent) insidious neck pain but conflict with other studies that have demonstrated impaired head repositioning accuracy (Kristjansson et al., 2003; Sjolander et al., 2008) and improvements in KS after cervical spine therapies (Palmgren et al., 2006). All previous studies utilised repositioning tasks to a NHP but as previously highlighted, variations in testing protocol, differences in measurement equipment and sample size may account for the differences in findings. It is also possible that the conflicting results arose from differences in patient populations such as pain severity and relative chronicity. The patient group in the earlier study in this thesis (Chapter 2) could be considered to be more comparable with those investigated by Teng et al., (2007)

When the peripheral proprioceptive signalling from the cervical spine is considered, it is possible that the tissue dysfunction, damage and inflammation associated with ‘whiplash’ injury could result in more obvious kinaesthetic impairments compared with neck pain of insidious onset. Although the cervical facet joint capsules contain a relatively high density and distribution of mechanoreceptors (McLain, 1994). It is the small intrinsic muscles, (particularly deep suboccipital muscles) which are likely to play a primary role in signalling the cervical proprioceptive information involved in the conscious perception of equilibrium, position and spatial orientation. (Nitz and Peck, 1986; Proske et al., 1988; Wilson, 1992; McLain, 1994; Bolton, 1998). One explanation that has been offered for the diminished kinaesthesia found, in groups of chronic neck pain patients, involves a functional alteration

in the muscle spindle receptors (Revel et al., 1991). This functional deficit could occur as a result of muscle pain (Pedersen et al., 1992; Matre, Sinkjaer, Svensson and Arendt-Nielsen, 1998) but importantly also from articular pain and dysfunction (Schaible and Grubb, 1993; Hurley, Jones and Newham, 1994; Hurley, Scott, Rees and Newham, 1997). The cervical facet joints have been well documented as a source of nociception in chronic neck pain particularly after a cervical injury such as 'whiplash' (Bogduk and Marsland, 1988; Aprill and Bogduk, 1992; Barnsley, Lord, Wallis and Bogduk, 1995; Lord, Barnsley, Wallis and Bogduk, 1996; Ketroser, 2000). However, it has been suggested that atrophy and fatty infiltration in the deep suboccipital muscles may lead to diminished or altered proprioceptive input to higher centres (Hallgren, Greenman and Rechten, 1994; McPartland and Brodeur, 1999), as evidenced by a reduced standing balance performance (McPartland, Brodeur and Hallgren, 1997). These findings again may well be the result of chronic nociception, inhibition from articular dysfunction or simply from disuse. However, in a patient with chronic neck pain and suboccipital atrophy after a forced flexion cervical injury, electromyography and magnetic resonance imaging abnormalities provided some evidence of denervation possibly as a result of nerve damage from trauma to the C1 dorsal ramus (Andary, Hallgren, Greenman and Rechten, 1998).

The existence of muscle pain in spinal pain syndromes is controversial (Bogduk, 1995). With respect to chronic neck pain of traumatic and non-traumatic origins, relatively little attention has been given to muscle tissue as a source of pain and dysfunction. There is some evidence to suggest that there are overlaps in the underlying muscle pathology and pathophysiology in different aetiologies of chronic neck pain (Weber, Uhlig, Grob, Dvorak and Muntener, 1993; Uhlig, Weber, Grob and Muntener, 1995). However, with regard to 'whiplash' injury, in addition to any possible functional disturbances in muscle mechanoreception because of muscle pain and or facet/facet capsular insult, the potential does exist for extrafusal muscle fibre damage (Brault, Siegmund and Wheeler, 2000). Although there is no direct evidence, it is therefore reasonable to suspect that this potential damage and inflammation may also extend to include damage and alteration in function of the muscle spindles and associated receptor endings. Although theoretically possible, the limited research has yet to demonstrate that 'whiplash' patients exhibit a significantly more pronounced cervicocephalic kinaesthetic deficit than patients with insidious onset neck pain. The only studies that had compared these patient groups directly before the current study commenced, did find a trend to suggest the kinaesthetic deficit may be greater in 'whiplash'

patients (Kristjansson et al., 2003). However, the reported pain and disability of the ‘whiplash’ group was also higher, which may have contributed to this difference. A recent study has also found greater repositioning error with both insidious neck pain and ‘whiplash’ groups compared with controls with a trend for the greatest errors with the ‘whiplash’ subjects (Sjolander et al., 2008). Although the NDI score was higher with the ‘whiplash’ group, the mean pain intensity and duration were higher with the insidious onset neck pain group. The question therefore still remains as to whether patients with insidious onset neck pain exhibit deficits in cervicocephalic sensibility or whether it is predominantly and more consistently seen with patients who have specifically suffered a ‘whiplash’ type injury at onset (Heikkila and Astrom, 1996; Loudon et al., 1997; Heikkila and Wenngren, 1998; Sterling et al., 2003; Treleaven et al., 2003 ; Kristjansson et al., 2004; Feipel et al., 2006; Treleaven et al., 2008).

The purpose of this study was to further investigate head repositioning accuracy in patients with chronic insidious onset neck pain but on this occasion using the Zebris CMS 70P system and three different testing protocols. These were two tests of HRA-SSA and a test of HRA to set angles within the horizontal plane of motion (left and right rotation). This non-straight ahead test was introduced to see if it might better detect errors in repositioning accuracy (Kristjansson et al., 2003). Kristjansson et al., (2003) utilised examiner positioned non-straight ahead target angles (30° left and right rotation) to investigate HRA but did not show a statistically significant difference between healthy subjects and insidious and traumatic onset patient groups although the insidious onset patients did show a slightly greater mean error than the control subjects (Table 2 – Chapter 1) A more recent study did reveal a significantly worse head repositioning for ‘whiplash’ patients compared with control subjects with a HRA task to 50° pure axial rotation and a combined rotation and lateral flexion position (Feipel et al., 2006). The current study modified Kristjansson’s (2003) protocol utilising two patient determined target angles in left and right rotation instead of one angle that was positioned by the examiner. The HRA-SSA tasks included the protocol used in the previous method agreement and *reliability* study (Chapter 5) and one using a protocol very similar to that used by Kristjansson et al., (2003), a testing protocol more closely related to that introduced by Revel et al., (1991). A ‘whiplash’ patient group was also included in this study to determine whether there were any differences related to the history of onset of neck pain.

## **6.2 METHODS**

### **6.2.1 Study Setting and Design**

This study took place in the outpatient clinic at the AECC, Bournemouth, UK. A prospective, three-group, observational cohort design was used. Completion of questionnaires and all measurement procedures were conducted in one of the radiography rooms. The same room was used on each occasion. Ethical approval was obtained from the AECC research ethics committee and written consent was obtained from all subjects before entering the study (Appendix V).

### **6.2.2 Subject Recruitment and Selection**

Male and female patients were initially selected as described in Chapter 2 (methods section) but over a six-month period. After the initial screening, those patients who met the general inclusion/exclusion criteria (Table 54) were provisionally subgrouped according to neck pain aetiology; cervical spine trauma (original and current neck pain resulting from cervical spine injury/trauma) or insidious (no history of cervical spine trauma). Subjects were excluded from participating in the study if the clinical records clearly showed that they had experienced cervical spine trauma/injury but the original or current onset of neck pain did not directly result from this. The remaining eligible and consenting patients were then contacted by phone and after a brief explanation of the nature of the study, invited to participate.

Those patients willing to take part were given an appointment with the investigator prior to their first visit after the new patient examination. At this measurement session with the investigator, the patients were given further verbal and written information about the study and asked to read and sign a consent form (Appendix V). Control subjects were recruited as described in Chapter 2 (methods section) using the inclusion/exclusion criteria shown in Table 54 from AECC staff, faculty and students. Subjects were invited to participate after an initial verbal presentation.

**Table 54.** General inclusion & exclusion criteria for both neck pain patient groups

---

Inclusion

1. Age 18-55.
2. Males and females.
3. Continuous neck pain of more than 7 weeks duration.
4. Neck pain as the only current painful musculoskeletal complaint.

Exclusion

1. Cervical Radiculopathy &/or Myelopathy
  2. Inflammatory Arthritis involving C-spine
  3. Tumour or infection involving C-spine
  4. Vertebrobasilar Artery Insufficiency
  5. Neurological disease such as MS, MND, Parkinson's, Syringomyelia etc.
  6. History of Dizziness
  7. Known congenital anomalies involving the C-spine.
  8. Systemic disease such as Diabetes Mellitus.
- 

*Abbreviations:*

C-Spine – Cervical Spine

MS – Multiple Sclerosis

MND – Motor Neurone Disease

---

### 6.2.3 Outcome Measures and Instrumentation

*Clinical characteristics of the neck pain groups.* These were collected as described in Chapter 2. Further, more detailed clinical data regarding the aetiology and temporal profile of the patients' neck pain, was obtained using a simple questionnaire generated specifically for this study by the investigator (Appendix V).

*Kinaesthetic sensibility tests.* Measurement of cervicocephalic kinaesthetic sensibility was performed using Zebris CMS 70P system as described in the last study (Chapter 5 – methods section).

### 6.2.4 Measurement protocol

The same examiner (main investigator) conducted all measurement procedures. On arrival, all participants were given an information sheet to read following which they were asked to sign the consent form (Appendix V). The neck pain subjects were next asked to complete the outcome questionnaires. All subjects were then checked again with respect to their

inclusion/exclusion criteria and questioned about any recent/new factors that may affect kinaesthetic performance.

For each of the three head repositioning tests, the subjects were seated and the Zebris equipment fitted as described in Chapter 5 (methods section). However, a laser pointer target was not used in this study and the room lights were switched off. No feedback on performance was given during the testing session.

Three testing protocols were investigated in this study; 1) head repositioning accuracy to the subjective straight ahead position (HRA-SSA), 2) HRA-SSA with feedback and 3) head positioning and repositioning accuracy to subjective non-straight ahead positions (HPA & HRA-nSA). The sequence of testing protocols was varied between subjects to minimize the possible effect of order on between test comparisons. The subjects were given a short rest (approx., 2 mins) between testing procedures during which the blindfold was raised and the next specific testing procedure was explained.

***Head repositioning accuracy to the subjective straight ahead position (HRA-SSA).*** The testing procedure for HRA-SSA was the same as used in the previous study (Chapter 5 – methods section) but without the laser pointer and target. A 60 seconds rest was allowed & the blindfold was not removed between trials..

***Head repositioning accuracy to SSA with feedback.*** The protocol for this procedure was identical to that for HRA-SSA other than proprioceptive feedback was given to each subject. After each repetition of HRA-SSA, the examiner gently relocated the subjects head to the zero position (if error was apparent) using the laser pointer as the guide. After this repositioning by the examiner, the subjects were given 2-3 seconds to concentrate on this feedback before the next movement instruction was given. Only two repositioning trials of 5 repetitions around the  $\theta Y$  axis of rotation (LR and RR  $\Rightarrow$  0) were performed by each subject. The sequence of trials was varied between subjects to minimize the possible effect of order on between trial comparisons.

***Head repositioning accuracy to nSA.*** For this testing procedure, the subjects were asked to position and then reposition the head to 30° and 45° left and right rotation angles from the SSA starting position. Before commencing the testing the subjects were reminded of the



size of a 30° and 45° angle using a diagram together with an explanation of the procedure in general. Once blindfolded, the subject was first asked to find what they perceived as the 'straight ahead' (SSA) position of their head with respect to the trunk as previously defined and notify the examiner verbally when they felt correctly aligned. The Zebris system was then quickly set to record mode, 'zeroed' (calibrated) so that this position was also defined as 0° for this equipment. The subject was then asked to recognise and remember this position for 2-3 seconds as this would be the reference they would need to use to judge the size of angular displacement. They were also reminded that this was the SSA position that they should try to relocate back to as accurately as possible after each non-straight ahead repositioning task. To initiate the repositioning task (as described earlier), the subject was then simply instructed to move the head to the target angle in the direction requested using the pre-recorded movement instructions (MPEG files). Once the subject had appeared to stop moving at the nSA position, the examiner waited approximately 6-8 seconds before giving the instruction to move back to the SSA position. Once again, after approximately 6-8 seconds, the subject was asked to reposition back to the same NSA angle. Five repetitions of HRA to the nSA position were undertaken with each trial direction. Four repositioning trials of left and right rotation to 30 and 45 degrees were performed. A new SSA position was established for each trial after approximately 60 seconds rest between each trial. The blindfold was not removed between trials. The sequence of trials was varied between subjects to minimize the possible effect of order on between trial comparisons.

### **6.2.5 Data Processing**

The raw angular 3-D position data for each trial, processed and saved by the Zebris system was processed as described in Chapter 5 (method section).

### **6.2.6 Data Analysis**

The absolute (unsigned) error (AE) and the signed error (SE) were used for comparative analysis of HRA for each measurement method. The AE was derived from the original directional data (SE) as described in Chapter 3 (methods section). The AE and SE for analysis of the non-straight ahead tasks were derived by treating the target angles of 30° and 45° as a zero degree position and then recalculating the error from this new mathematical point. For example, a head position of 48° left rotation from the SSA position would be recalculated as a SE of -3° and an AE of 3° from a target of 45° left rotation (-45° using the

orthogonal coordinates). This was done to allow comparison of HRA between the straight ahead and non-straight ahead tasks. The HRA data and all other interval data for analysis (e.g., neck pain severity and duration, age and head movement characteristics) were examined for normality using a combination of statistical testing (Kolmogorov-Smirnov or K-S test) and observation of the data using frequency histograms, Q-Q (normal) plots and box plot summaries of the measures of central tendency (mean and median) and distribution. Following this analysis, all interval data were considered normal in distribution and therefore parametric statistics were used for analyses of these data.

Differences in the normally distributed interval data were analysed with one-way independent ANOVA (3 group comparisons) or unpaired *t*-test (2 group comparisons) after testing for homogeneity of variance using Levene's test. Where unequal variances were demonstrated, the one-way independent ANOVA or unpaired *t*-test was performed with Welch correction. To test for differences between 'pairs' of groups when variances were equal, Gabriel's *post hoc* procedure was used. When the assumption of homogeneity of variance was in doubt or clearly violated, the Games-Howell procedure was utilised as the *post hoc* test. Differences between groups for nominal variables (e.g., gender, location of neck pain etc) were tested using Pearson's chi-square test of independence.

The results have been presented using the guidelines produced by Lang and Secic (1997) and Field (2005) and all statistical analyses were performed using SPSS for Windows.

## **6.3 RESULTS**

### **6.3.1 Group demographics and clinical characteristics**

Over the six-month study period, thirty-four new patients to the AECC teaching clinic satisfied the inclusion/exclusion criteria and were willing and able to participate. Eighteen patients were classified with neck pain related to trauma (NP-T) and sixteen with neck pain of an insidious nature (NP-I). Seventeen volunteers were recruited as the control subjects. The group demographics (age and gender) and neck pain characteristics of the cervicogenic groups are shown in Table 55.

**Table 55.** Group demographics and clinical characteristics.

Groups	n	Age (years)		Gender (M:F)	Neck Pain (NP) Characteristics	
		Mean ± SD	Range		NP Severity* (Mean ± SD)	Total NP Duration (weeks)
Healthy	17	33.6 ± 10.48	19 - 53	7:10	-	-
NP - Insidious	16	31.3 ± 8.27	23 - 52	5:11	4.1 ± 1.7	172.2 ± 216.10
NP - Trauma	18	37.5 ± 8.63	18 - 53	8:10	5.1 ± 1.6	243.3 ± 266.71

Abbreviations: NRS – numerical rating scale

\* Average NP severity over the 2-3 days immediately prior to testing; measured using numerical rating scale (NRS).

Analysis of the demographic data did not indicate any statistically significant differences among the groups in mean ages with one-way independent ANOVA ( $F_{(2,48)} = 0.94$ ;  $P = 0.40$ ) and gender distribution with Chi-square test ( $\chi^2(2) = 0.66$ ;  $P = 0.99$ ). A comparison of the neck pain characteristics showed that the neck pain severity and total duration were higher in the NP-T group. However, analysis with an unpaired  $t$ -test indicated that these differences were not statistically significant (NP duration -  $d = 71.1$ ; 95% CI = -99.9 to 242.0;  $t = 0.85$ ;  $df = 32$ ;  $P = 0.40$ : NP severity -  $d = 1.0$ ; 95% CI = -0.14 to 2.13;  $t = 1.79$ ;  $df = 32$ ;  $P = 0.08$ ).

### 6.3.2 Kinaesthetic Sensibility Testing

#### Test 1 - Head repositioning accuracy to subject straight ahead (HRA-SSA)

A graphical overview of the individual subject HRA-SSA (mean signed values of five repositioning repetitions) for the patient and control groups, with each of the four movement directions trials, is presented in Figure 54 using scatter plots. Only the  $\theta_x$  (vertical) and  $\theta_y$  (horizontal) components are shown in these plots. A visual examination of the scatterplots does not show an obvious difference in cluster pattern between the groups other than a relative overshoot characteristic for NP-T group with the Ext  $\Rightarrow$  0 trial.

**Head repositioning accuracy (HRA-SSA): absolute (unsigned) values.** The repositioning errors for the NP-T group were generally slightly higher in most (6 of 8) movement direction components compared with the other two groups (Table 56). However, in most

cases, this difference was no more than 0.5 degrees. Analysis with one-way independent ANOVA indicated a statistically significant difference among the groups for the  $\theta Y$  component of the RR  $\Rightarrow$  0 trial only (Table 56). Further analysis with *post hoc* pair-wise comparisons revealed a statistically significant difference between the NP-I and NP-T groups using Gabriel's procedure ( $d = 1.8$ ; 95% CI = 0.09 to 3.49;  $P < 0.05$ ). However, this difference was not significant with the Games-Howell *post hoc* procedure ( $d = 1.8$ ; 95% CI = -0.38 to 3.62;  $P = 0.056$ ). No significant difference was seen between the symptomatic groups and the control subjects. It is perhaps worth noting that for all groups, the absolute repositioning errors around the axis perpendicular to the primary motion were relatively smaller for the vertical movement trials (Ext and Flex  $\Rightarrow$  0) than for the horizontal movement trial (LR and RR  $\Rightarrow$  0). It is also interesting to note that with the exception of the NP-T group, the  $\theta X$  absolute repositioning error (primary motion direction) for the Ext  $\Rightarrow$  0 trial was the largest of all the error components observed (Table 56).

**Table 56.** Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-SSA in insidious ( $n=16$ ) and traumatic ( $n=18$ ) onset neck pain patients and healthy subjects ( $n=17$ ).

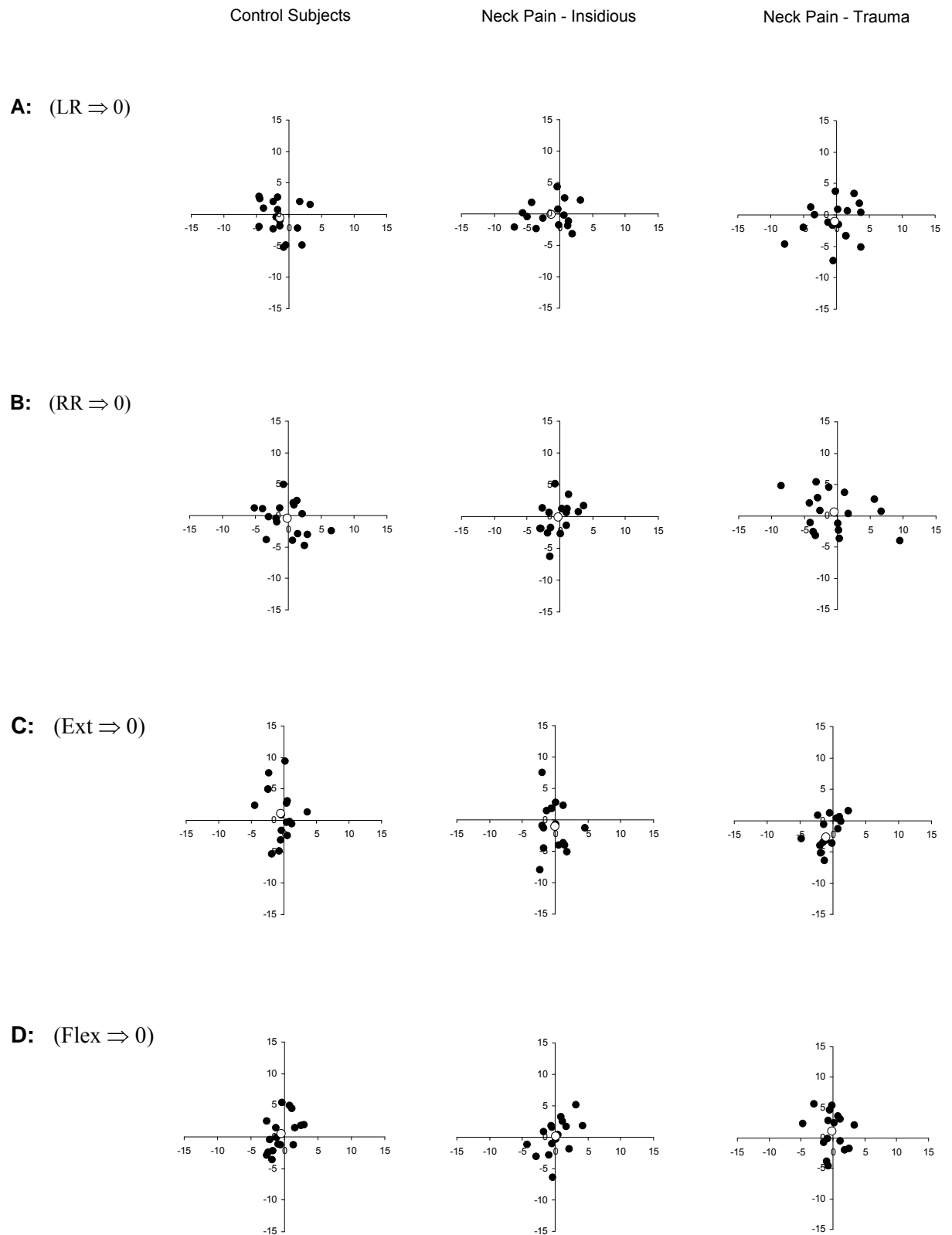
Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR	$\theta X$	2.4 $\pm$ 1.45	1.62 $\rightarrow$ 3.10	1.6 $\pm$ 1.18	0.96 $\rightarrow$ 2.22	2.5 $\pm$ 1.96	1.48 $\rightarrow$ 3.50	2,47	1.56	0.22
	$\theta Y^\dagger$	2.3 $\pm$ 1.33	1.66 $\rightarrow$ 3.03	2.4 $\pm$ 2.08	1.32 $\rightarrow$ 3.54	2.5 $\pm$ 2.08	1.38 $\rightarrow$ 3.52	2,47	0.02	0.99
RR	$\theta X$	2.2 $\pm$ 1.53	1.41 $\rightarrow$ 2.99	2.2 $\pm$ 1.61	1.31 $\rightarrow$ 3.02	2.7 $\pm$ 1.53	1.90 $\rightarrow$ 3.48	2,47	0.60	0.55
	$\theta Y^\dagger$	2.4 $\pm$ 1.64	1.52 $\rightarrow$ 3.20	1.7 $\pm$ 1.02	1.14 $\rightarrow$ 2.23	3.5 $\pm$ 2.79	2.04 $\rightarrow$ 4.91	2,47	3.45**	<0.05**
Ext	$\theta X^\dagger$	3.3 $\pm$ 2.58	1.94 $\rightarrow$ 4.60	3.2 $\pm$ 2.24	2.00 $\rightarrow$ 4.40	3.3 $\pm$ 4.06	1.08 $\rightarrow$ 5.42	2,46	0.00	1.00
	$\theta Y$	1.4 $\pm$ 1.26	0.72 $\rightarrow$ 2.02	1.4 $\pm$ 1.07	0.87 $\rightarrow$ 2.01	1.8 $\pm$ 1.69	0.93 $\rightarrow$ 2.73	2,46	0.54	0.59
Flex	$\theta X^\dagger$	2.3 $\pm$ 1.53	1.52 $\rightarrow$ 3.09	2.2 $\pm$ 1.65	1.36 $\rightarrow$ 3.11	2.9 $\pm$ 1.66	1.97 $\rightarrow$ 3.74	2,46	0.72	0.49
	$\theta Y$	1.6 $\pm$ 0.78	1.24 $\rightarrow$ 2.04	1.6 $\pm$ 1.37	0.86 $\rightarrow$ 2.32	1.5 $\pm$ 1.28	0.82 $\rightarrow$ 2.19	2,46	0.05	0.95

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex).

$\dagger$  Primary axis of rotation for the repositioning test

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis. A statistically significant result is represented by  $P < 0.05$

\*\*Indicates F-ratio & P-value with Welch Correction.



**Figure 54.** Scatterplots showing the head repositioning accuracy (HRA) for the neck pain patients and control subjects. All data are presented as degrees. Abbreviations: HRA to reference zero following a near-maximal active head movement to reference zero; left rotation (LR ⇒ 0) , right rotation (RR ⇒ 0), extension (Ext ⇒ 0) and flexion (Flex ⇒ 0); The empty circle ( O ) on the scatterplots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).

**Head repositioning accuracy (HRA-SSA): signed values.** The mean ( $\pm$  1SD) signed repositioning errors around the primary axis of rotation for each group are presented in Table 57. This index of HRA is also represented graphically by the empty circle on the scatterplots ( O ) shown in Figures 54 and 55 using box plots. Analysis with one-way independent ANOVA indicated a statistically significant difference among the groups for the Ext  $\Rightarrow$  0 trial only (Table 57). Further analysis of this difference with *post hoc* pair-wise comparisons revealed a statistically significant undershoot in the NP-T group compared to healthy subjects using Gabriel's procedure ( $d = 3.8$ ; 95% CI = 0.25 to 7.38;  $P < 0.05$ ) and the Games-Howell procedure ( $d = 3.8$ ; 95% CI = 0.14 to 7.48;  $P < 0.05$ ).

**Table 57.** Comparison of signed repositioning error (degrees) for HRA-SSA in insidious ( $n=16$ ) and traumatic ( $n=18$ ) onset neck pain patients and healthy subjects ( $n=17$ ) in the primary motion direction.

Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR	$\theta Y^\dagger$	-1.4 $\pm$ 2.38	-2.58 $\rightarrow$ -0.13	-1.3 $\pm$ 2.99	-2.87 $\rightarrow$ 0.32	-0.2 $\pm$ 3.26	-1.87 $\rightarrow$ 1.49	2,47	0.85	0.43
RR	$\theta Y^\dagger$	-0.1 $\pm$ 2.93	-1.58 $\rightarrow$ 1.44	-0.3 $\pm$ 2.00	-1.32 $\rightarrow$ 0.82	-0.5 $\pm$ 4.51	-2.82 $\rightarrow$ 1.82	2,47	0.06**	0.95**
Ext	$\theta X^\dagger$	1.1 $\pm$ 4.10	-1.04 $\rightarrow$ 3.18	-1.0 $\pm$ 3.86	-3.07 $\rightarrow$ 1.04	-2.7 $\pm$ 4.45	-5.11 $\rightarrow$ -0.38	2,46	3.51	<0.05
Flex	$\theta X^\dagger$	0.5 $\pm$ 2.78	-0.94 $\rightarrow$ 1.92	0.1 $\pm$ 2.83	-1.41 $\rightarrow$ 1.61	1.1 $\pm$ 3.17	-0.54 $\rightarrow$ 2.84	2,46	0.52	0.60

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex).

$^\dagger$  Primary axis of rotation for the repositioning test

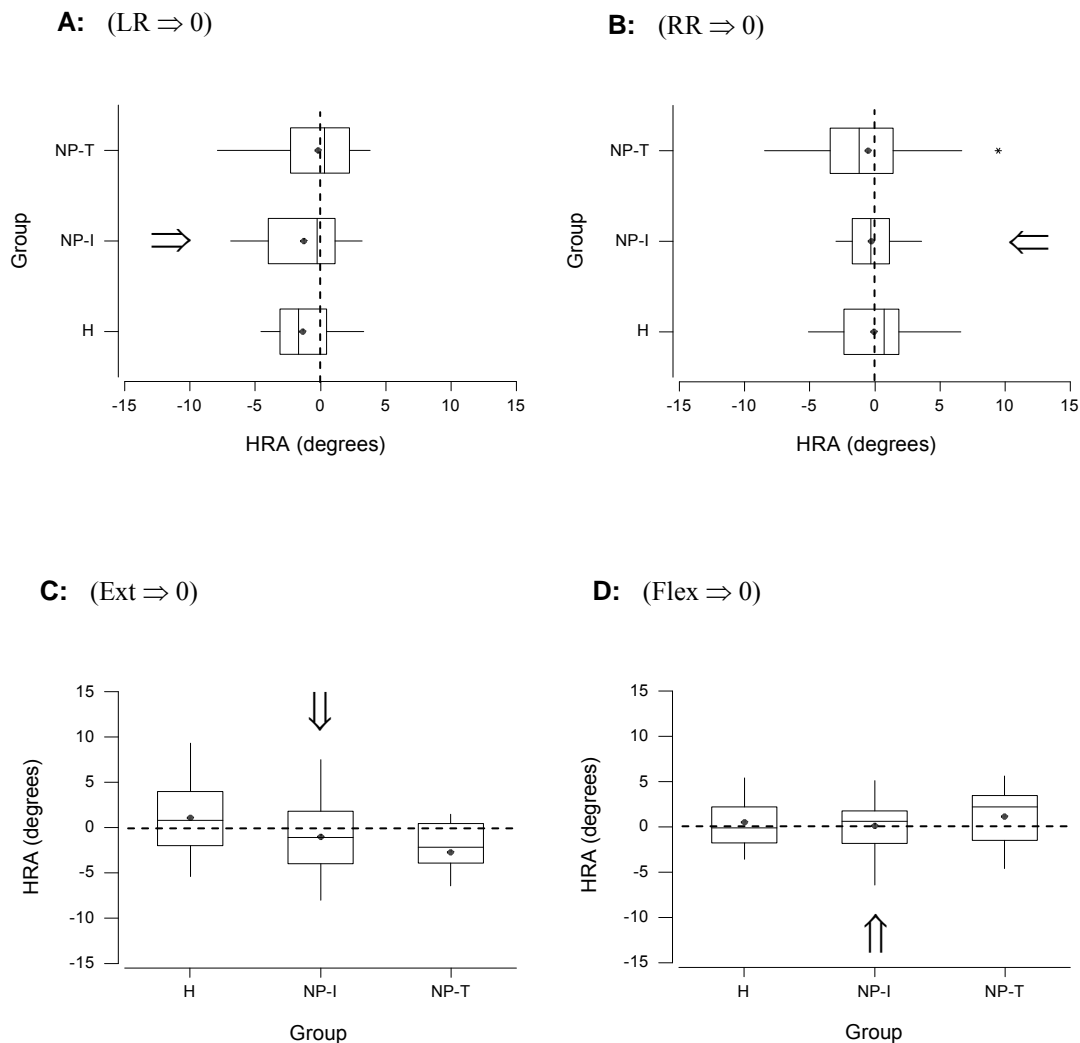
\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis. A statistically significant result is represented by  $P < 0.05$

\*\*Indicates F-ratio & P-value with Welch Correction.

### **Head Movement Velocity and Range of Motion during kinaesthetic sensibility testing.**

The mean head movement velocities (both away and towards the repositioning target) for the NP-T and NP-I groups were generally reduced in all trials compared with the control group (Table 58). Analysis with one-way independent ANOVA indicated a statistically significant difference among the groups only for the Ext  $\Rightarrow$  0 trial (Table 58). Further analysis of this using Gabriel's *post hoc* procedure revealed the difference was statistically significant between the healthy subject and trauma NP-T groups for the outward movement velocity ( $d = 10.6$ ; 95% CI = 1.51 to 19.83;  $P < 0.05$ ), the return to target (SSA) movement velocity ( $d = -8.4$ ; 95% CI = -17.10 to 0.37;  $P < 0.05$ ). The NP-T group also showed a

slight reduction in the peak active cervical range of motion (ROM) during the tests compared with the other two groups. Using one-way independent ANOVA, the difference was again only statistically significant among the groups for the Ext  $\Rightarrow$  0 trial. A *post hoc* pair-wise comparison with Gabriel's procedure demonstrated the difference was statistically significant between the healthy subject and trauma NP-T groups ( $d = 11.6$ ; 95% CI = 0.93 to 22.34;  $P < 0.05$ ).



**Figure 55.** Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H). The horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) are displayed for L & RR  $\Rightarrow$  0 (A-B) and Ext & Flex  $\Rightarrow$  0 (C-D) respectively. The middle horizontal/vertical bar represents the median value; the box bottom and top or left and right for L & RR, represent the interquartile range (25<sup>th</sup> and 75<sup>th</sup> percentile); the mean value is represented by the dot.; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:Q1 - 1.5 (Q3 - Q1) Upper Limit:Q1 + 1.5 (Q3 - Q1)).

**Table 58.** Comparison of movement characteristics during HRA-SSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction

Primary Motion	Movement Parameter	Healthy Subjects	Neck Pain (Insidious)	Neck Pain (Trauma)	df	F-test	P-Value*
		Mean ± SD	Mean ± SD	Mean ± SD			
LR	Velocity 1 (°/s)	-48.7 ± 14.69	-39.0 ± 12.41	-39.8 ± 19.56	2,31	2.27**	0.12**
	Velocity 2 (°/s)	44.5 ± 12.89	36.4 ± 8.55	35.7 ± 15.41	2,30	2.53**	0.10**
	Peak ROM (degrees)	-66.8 ± 9.01	-67.0 ± 9.34	-61.1 ± 13.59	2,31	1.24**	0.30**
RR	Velocity 1 (°/s)	48.9 ± 17.61	39.6 ± 11.77	40.3 ± 16.97	2,47	1.82	0.17
	Velocity 2 (°/s)	-44.0 ± 15.01	-36.4 ± 6.59	-38.3 ± 16.14	2,28	1.77**	0.19**
	Peak ROM (degrees)	66.3 ± 9.99	63.9 ± 9.91	61.7 ± 10.83	2,47	0.86	0.43
-----							
Ext	Velocity 1 (°/s)	37.3 ± 10.29	29.0 ± 10.04	26.6 ± 11.52	2,46	4.61	<0.05
	Velocity 2 (°/s)	-35.1 ± 11.32	-27.5 ± 7.94	-26.7 ± 10.76	2,46	3.45	<0.05
	Peak ROM (degrees)	59.4 ± 10.48	53.3 ± 11.41	47.8 ± 15.02	2,46	3.63	<0.05
Flex	Velocity 1 (°/s)	-35.8 ± 12.87	-29.4 ± 10.68	-26.8 ± 13.41	2,45	2.30	0.11
	Velocity 2 (°/s)	32.6 ± 10.53	26.4 ± 8.28	25.0 ± 11.22	2,28	2.79	0.07
	Peak ROM (degrees)	-54.1 ± 11.70	-54.2 ± 10.52	-46.8 ± 11.55	2,47	2.23	0.12

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); range of motion (ROM)

Velocity 1 – mean velocity away from target; Velocity 2 – mean velocity returning to target

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA. A statistically significant result is represented by  $P < 0.05$

\*\*Indicates F-ratio & P-value with Welch Correction.

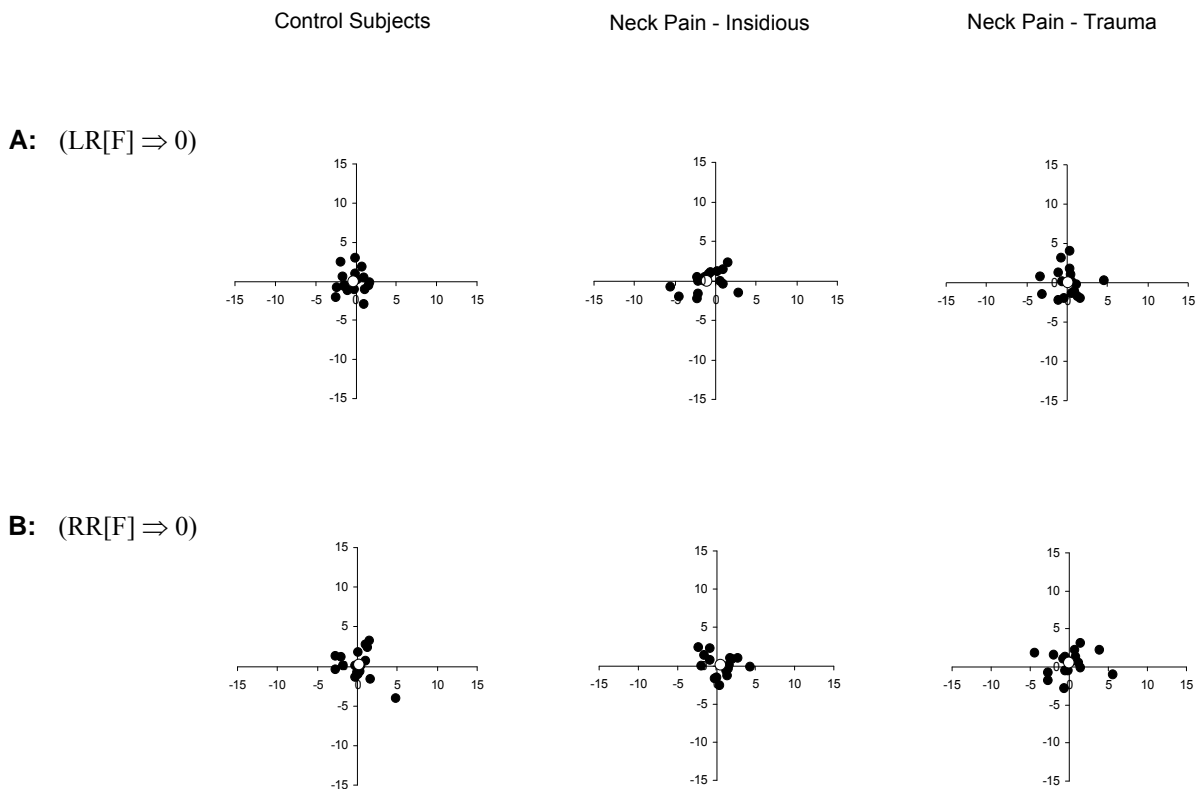
## Test 2 - Head repositioning accuracy to subjective straight ahead with feedback (HRA-SSA[F]).

A graphical overview of the individual subject HRA-SSA (mean values of five repositioning repetitions) for the patient and control groups, with each of the trials, is presented in Figure 56 using scatter plots. A visual examination of the scatterplots does not show a difference in cluster pattern between the groups

**Head repositioning accuracy (HRA-SSA) with feedback: absolute (unsigned) values.** The mean absolute (unsigned) horizontal ( $\theta Y$ ), and vertical ( $\theta X$ ) repositioning errors with each of the four movement directions trials are presented in (Table 59). Analysis with one-way



independent ANOVA did not indicate any statistically significant difference among the groups (Table 59). However, it is interesting to note that for all directional components in the horizontal rotation trials, the absolute repositioning errors and variance within each group were lower compared with the test performed without a feedback component (Table 56).



**Figure 56.** Scatterplots showing the head repositioning accuracy (HRA) with feedback for the neck pain patients and healthy subjects. All data are presented as degrees. Abbreviations: HRA to reference zero following a near-maximal active head movement to reference zero; left rotation (LR => 0) , right rotation (RR => 0); The empty circle ( O ) on the scatterplots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).

**Table 59.** Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-SSA with feedback, in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17).

Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR	$\theta X$	1.2 $\pm$ 0.91	0.74 $\rightarrow$ 1.68	1.0 $\pm$ 0.76	0.63 $\rightarrow$ 1.44	1.4 $\pm$ 1.07	0.90 $\rightarrow$ 2.00	2,47	0.82	0.45
	$\theta Y^\dagger$	1.2 $\pm$ 0.75	0.78 $\rightarrow$ 1.55	1.9 $\pm$ 1.44	1.12 $\rightarrow$ 2.66	1.3 $\pm$ 1.21	0.68 $\rightarrow$ 1.93	2,47	1.78	0.18
RR	$\theta X$	1.4 $\pm$ 1.14	0.79 $\rightarrow$ 1.97	1.1 $\pm$ 0.81	0.70 $\rightarrow$ 1.57	1.4 $\pm$ 0.86	0.93 $\rightarrow$ 1.81	2,47	0.36	0.70
	$\theta Y^\dagger$	1.3 $\pm$ 1.27	0.67 $\rightarrow$ 1.97	1.5 $\pm$ 1.06	0.96 $\rightarrow$ 2.10	1.7 $\pm$ 1.59	0.92 $\rightarrow$ 2.55	2,47	0.40	0.67

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex).

$^\dagger$  Primary axis of rotation for the repositioning test

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis. A statistically significant result is represented by  $P < 0.05$

\*\*Indicates F-ratio & P-value with Welch Correction.

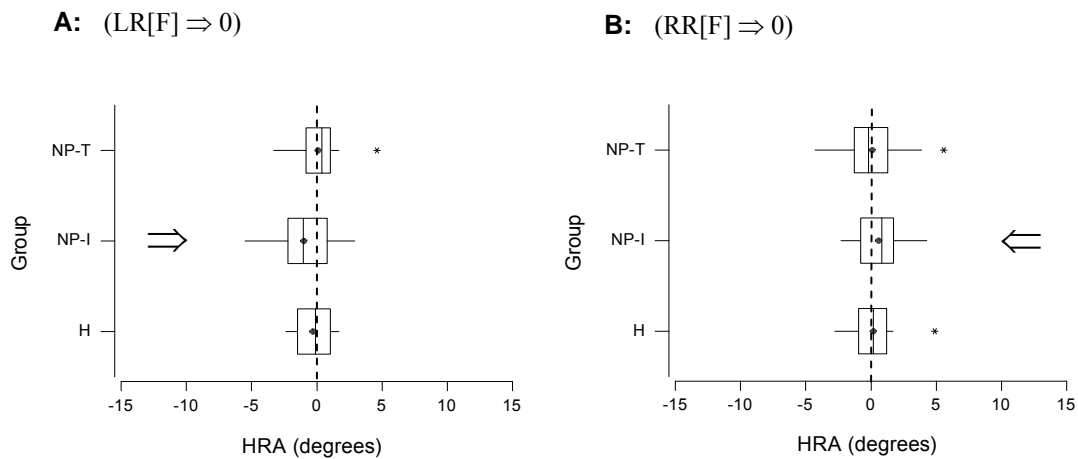
**Head repositioning accuracy (HRA-SSA) with feedback: signed values.** The mean signed repositioning errors around the primary axis of rotation are presented in Table 60. This index of HRA is also represented graphically by the empty circle on the scatterplots ( O ) shown in Figures 56 and 57 using box plots. Analysis with one-way independent ANOVA did not indicate any statistically significant difference among the groups (Table 60).

**Table 60.** Comparison of signed repositioning error (degrees) for HRA-SSA with feedback in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction.

Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR	$\theta Y$	-0.3 $\pm$ 1.39	-1.00 $\rightarrow$ 0.42	-1.0 $\pm$ 2.19	-2.17 $\rightarrow$ 0.16	0.1 $\pm$ 1.81	-0.82 $\rightarrow$ 1.04	2,47	1.58	0.22
RR	$\theta Y$	0.2 $\pm$ 1.85	-0.77 $\rightarrow$ 1.14	0.6 $\pm$ 1.80	-0.37 $\rightarrow$ 1.56	0.1 $\pm$ 2.39	-1.14 $\rightarrow$ 1.31	2,47	0.29	0.75

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex).

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis. A statistically significant result is represented by  $P < 0.05$



**Figure 57.** Box plots of the signed repositioning error representing over/undershoot for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H). The horizontal ( $\theta_Y$ ) repositioning error (degrees) are displayed for L & RR  $\Rightarrow$  0 (A-B). The middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25<sup>th</sup> and 75<sup>th</sup> percentile); the mean value is represented by the dot.; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:Q1 - 1.5 (Q3 - Q1) Upper Limit:Q1 + 1.5 (Q3 - Q1)).

### ***Head Movement Velocity and Range of Motion during kinaesthetic sensibility testing.***

The characteristics of head movement in the primary motion direction for each of the trials are presented in Table 61. The mean head movement velocities (both away and towards the repositioning target) and peak active cervical ROM were generally reduced in all trials for the NP-T and NP-I groups compared with the control group. No statistically significant differences were indicated among the groups after analysis with one-way independent ANOVA (Table 61).

**Table 61.** Comparison of movement characteristics during HRA-SSA with feedback in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction

Primary Motion	Movement Parameter	Healthy Subjects	Neck Pain (Insidious)	Neck Pain (Trauma)	df	F-test	P-Value*
		Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD			
LR	Velocity 1 (m/s <sup>2</sup> )	-47.3 $\pm$ 13.49	-38.3 $\pm$ 17.07	-41.4 $\pm$ 15.92	2,47	1.78	0.18
	Velocity 2 (m/s <sup>2</sup> )	39.4 $\pm$ 11.60	33.0 $\pm$ 11.80	35.3 $\pm$ 14.46	2,47	1.08	0.35
	Peak ROM (degrees)	-65.1 $\pm$ 6.89	-63.9 $\pm$ 12.62	-59.1 $\pm$ 13.06	2,28	1.24**	0.30**
RR	Velocity 1 (m/s <sup>2</sup> )	46.8 $\pm$ 14.39	37.4 $\pm$ 16.15	40.7 $\pm$ 17.65	2,47	1.46	0.24
	Velocity 2 (m/s <sup>2</sup> )	-41.2 $\pm$ 14.72	-33.0 $\pm$ 11.09	-36.2 $\pm$ 14.50	2,47	1.51	0.23
	Peak ROM (degrees)	62.9 $\pm$ 9.34	60.6 $\pm$ 9.89	59.4 $\pm$ 11.34	2,47	0.50	0.61

*Abbreviations:* left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); range of motion (ROM)

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA. A statistically significant result is represented by P < 0.05

\*\*Indicates F-ratio & P-value with Welch Correction.

### Test 3 - Head positioning and repositioning accuracy to non-straight ahead (HRA-nSA).

A graphical overview of the individual subject HRA-NSA (mean values of five repositioning repetitions) for the patient and control groups is presented in Figures 58 & 59 using scatter plots. Only the  $\theta X$  (vertical) and  $\theta Y$  (horizontal) components are shown in these plots. No clear difference in cluster pattern was seen between the groups on a visual examination of the scatterplots although for the right rotation trials, a relatively large variability in repositioning error can be seen for the patients groups compared with the control subjects.

**Head repositioning accuracy to non-straight ahead (HRA-nSA): absolute values.** Within each of the trials, the absolute repositioning error was noticeably larger in the plane of primary movement ( $\theta Y$ ) for all groups and was generally slightly greater in this plane for the patient (symptomatic) groups compared with the control group (Table 62). Analysis with one-way independent ANOVA did not demonstrate a statistically significant difference between the groups for any of the trials (Table 62).

**Table 62.** Comparison of absolute (unsigned) Horizontal ( $\theta Y$ ) and Vertical ( $\theta X$ ) repositioning error (degrees) for HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17).

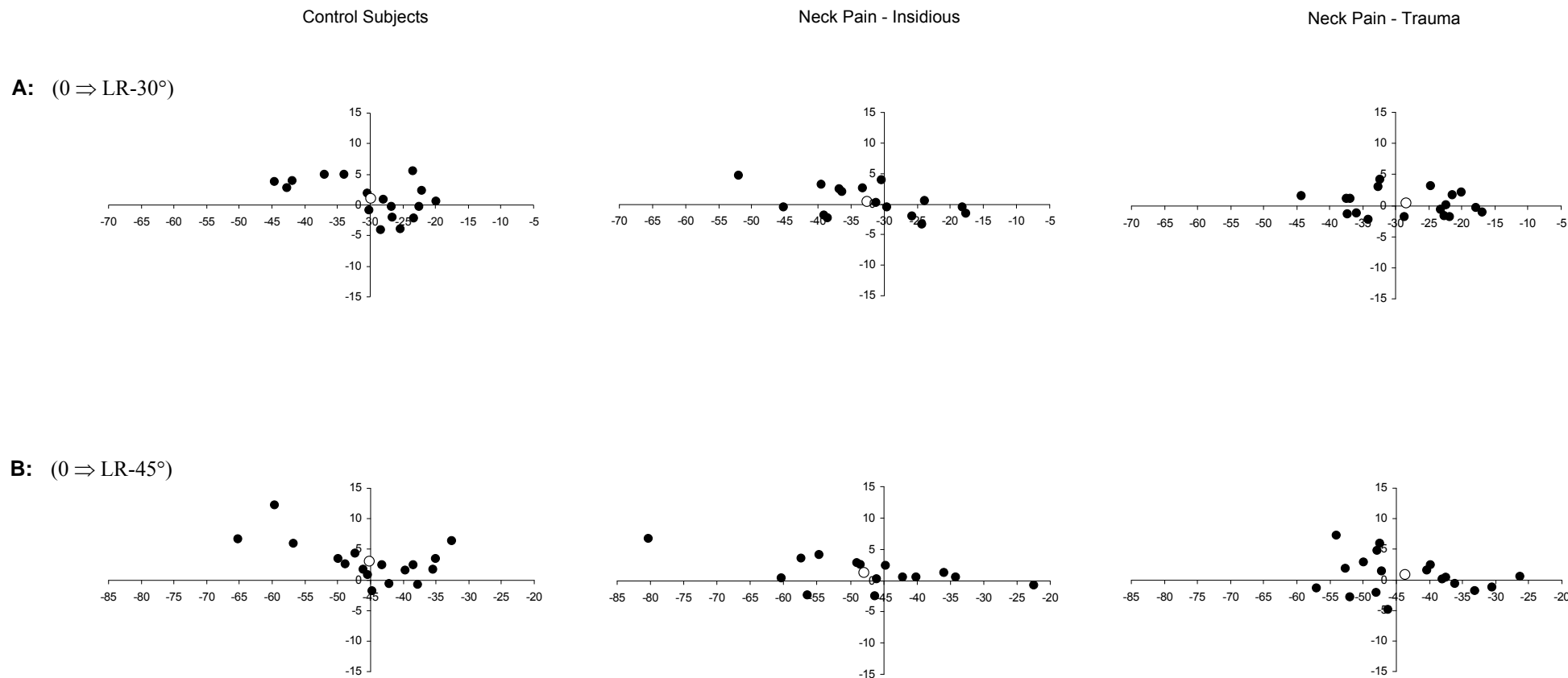
Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR-30	$\theta X$	2.7 $\pm$ 1.75	1.77 $\rightarrow$ 3.56	2.0 $\pm$ 1.33	1.30 $\rightarrow$ 2.72	1.7 $\pm$ 1.00	1.16 $\rightarrow$ 2.16	2,30	2.16**	0.13**
	$\theta Y-0^\dagger$	6.1 $\pm$ 4.33	3.90 $\rightarrow$ 8.35	7.7 $\pm$ 5.75	4.60 $\rightarrow$ 10.73	7.3 $\pm$ 3.54	5.55 $\rightarrow$ 9..07	2,48	0.52	0.60
LR-45	$\theta X$	3.5 $\pm$ 2.96	1.93 $\rightarrow$ 4.98	2.0 $\pm$ 1.78	1.06 $\rightarrow$ 3.03	2.4 $\pm$ 2.01	1.43 $\rightarrow$ 3.43	2,48	1.61	0.21
	$\theta Y-0^\dagger$	6.8 $\pm$ 5.61	3.89 $\rightarrow$ 9.66	9.6 $\pm$ 9.42	4.42 $\rightarrow$ 14.85	7.3 $\pm$ 4.70	4.92 $\rightarrow$ 9.59	2,48	0.82	0.45
RR-30	$\theta X$	2.1 $\pm$ 1.46	1.35 $\rightarrow$ 2.85	2.9 $\pm$ 2.24	1.69 $\rightarrow$ 4.08	2.2 $\pm$ 1.26	1.59 $\rightarrow$ 2.84	2,48	1.05	0.36
	$\theta Y-0^\dagger$	5.1 $\pm$ 5.15	2.48 $\rightarrow$ 7.78	5.5 $\pm$ 5.32	2.71 $\rightarrow$ 8.39	6.6 $\pm$ 4.68	4.24 $\rightarrow$ 8.89	2,48	0.38	0.69
RR-45	$\theta X$	2.6 $\pm$ 2.23	1.46 $\rightarrow$ 3.76	2.9 $\pm$ 2.83	1.36 $\rightarrow$ 4.49	2.6 $\pm$ 2.20	1.45 $\rightarrow$ 3.71	2,48	0.10	0.91
	$\theta Y-0^\dagger$	4.7 $\pm$ 7.50	0.85 $\rightarrow$ 8.57	8.9 $\pm$ 7.23	4.89 $\rightarrow$ 12.90	6.8 $\pm$ 4.57	4.47 $\rightarrow$ 9.16	2,48	1.63	0.21

*Abbreviations:* LR  $\Rightarrow$  30, RR  $\Rightarrow$  30; LR  $\Rightarrow$  45, RR  $\Rightarrow$  45; positioning to 30° left & right and 45° left & right rotation:  $\theta Y-0$  represents the error compared with the NSA target angle (e.g. 30° or 45°) calibrated as 0°.

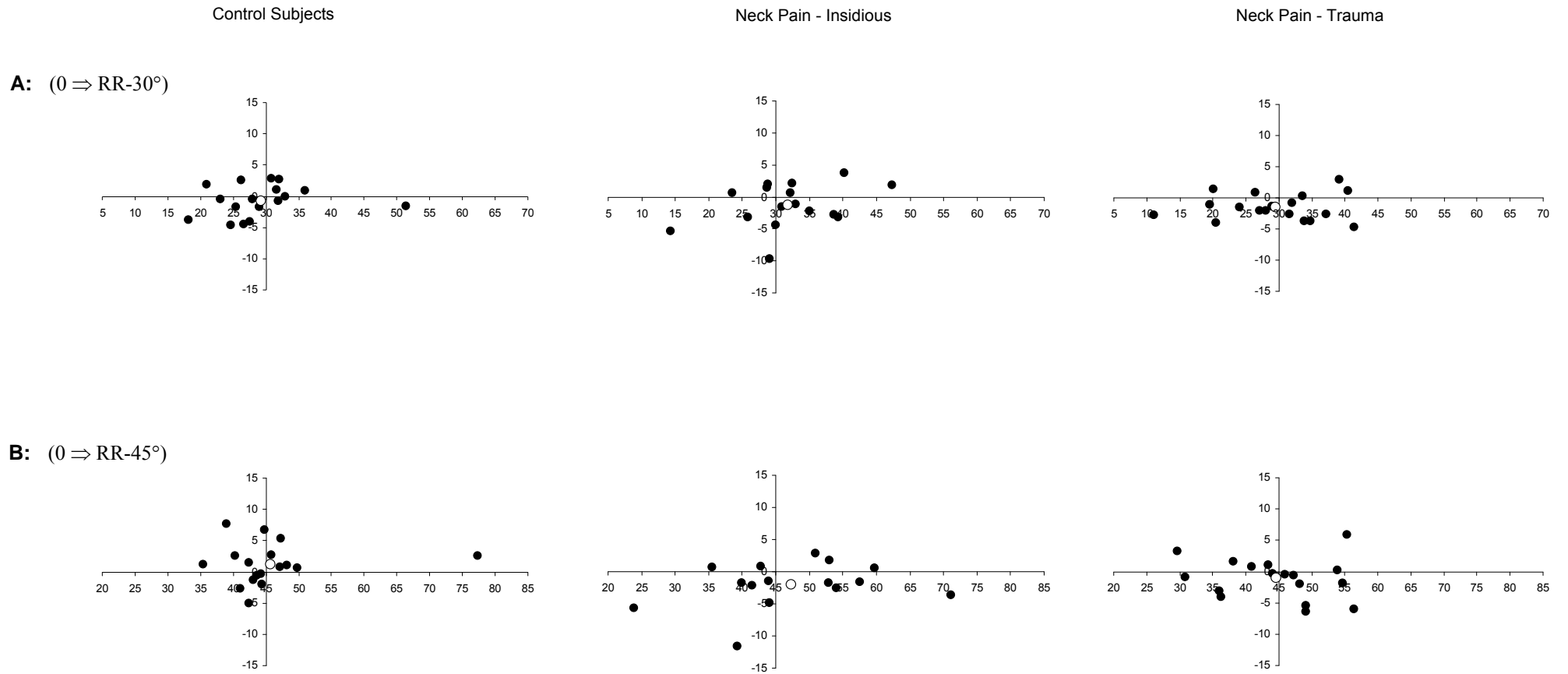
$\dagger$  Primary axis of rotation for the repositioning test

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis.

\*\*Indicates F-ratio & P-value with Welch Correction.



**Figure 58.** Scatterplots showing the head positioning accuracy to non-straight ahead targets (HPA-NSA) for the neck pain patients and healthy subjects. All data are presented as degrees. Abbreviations: HPA-NSA to targets 30 or 45 degrees from subjective straight ahead; positioning to 30° & 45° left rotation (LR ⇒ 30, LR ⇒ 45); The empty circle (O) on the scatterplots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement).



**Figure 59.** Scatterplots showing the head positioning accuracy to non straight ahead targets (HPA-NSA) for the neck pain patients and healthy subjects. All data are presented as degrees. Abbreviations: HPA-NSA to targets 30 or 45 degrees from subjective straight ahead; positioning to 30° & 45° right rotation (RR ⇒ 30, RR ⇒ 45); The empty circle (O) on the scatterplots is the central point of the data (represents overshoot/undershoot characteristics in plane of movement)

**Head positioning accuracy to non-straight ahead (HPA-NSA): signed values.** The mean ( $\pm$  1SD) signed repositioning errors around the primary axis of rotation for each group are presented in Table 63. This index of HRA is also represented graphically by the empty circle on the scatterplots ( O ) shown in Figures 58 and 59 and in Figure 60 using box plots. Analysis with one-way independent ANOVA did not demonstrate any statistically significant difference among the groups although a slight overshoot tendency was seen for the insidious onset neck pain groups compared with the control and ‘whiplash’ groups (Table 63).

**Table 63.** Comparison of signed repositioning error (degrees) for HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction.

Primary Motion	Axis of rotation	Healthy Subjects		Neck Pain (Insidious)		Neck Pain (Trauma)		df	F test	P-Value*
		Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI	Mean $\pm$ SD	95%CI			
LR-30	$\theta Y-0$	0.2 $\pm$ 7.65	-3.76 $\rightarrow$ 4.11	-2.6 $\pm$ 9.41	-7.61 $\rightarrow$ 2.41	1.6 $\pm$ 8.14	-2.42 $\rightarrow$ 5.67	2,48	1.10	0.34
LR-45	$\theta Y-0$	-0.2 $\pm$ 8.95	-4.77 $\rightarrow$ 4.34	-2.9 $\pm$ 13.39	-10.28 $\rightarrow$ 4.55	1.4 $\pm$ 8.69	-2.88 $\rightarrow$ 5.77	2,47	0.71	0.50
RR-30	$\theta Y-0$	-0.8 $\pm$ 7.33	-4.57 $\rightarrow$ 2.98	1.9 $\pm$ 7.59	-2.19 $\rightarrow$ 5.89	-0.6 $\pm$ 8.20	-4.64 $\rightarrow$ 3.52	2,48	0.59	0.56
RR-45	$\theta Y-0$	0.7 $\pm$ 8.91	-3.91 $\rightarrow$ 5.25	2.4 $\pm$ 11.45	-3.98 $\rightarrow$ 8.70	-0.3 $\pm$ 8.37	-4.64 $\rightarrow$ 3.97	2,48	0.32	0.73

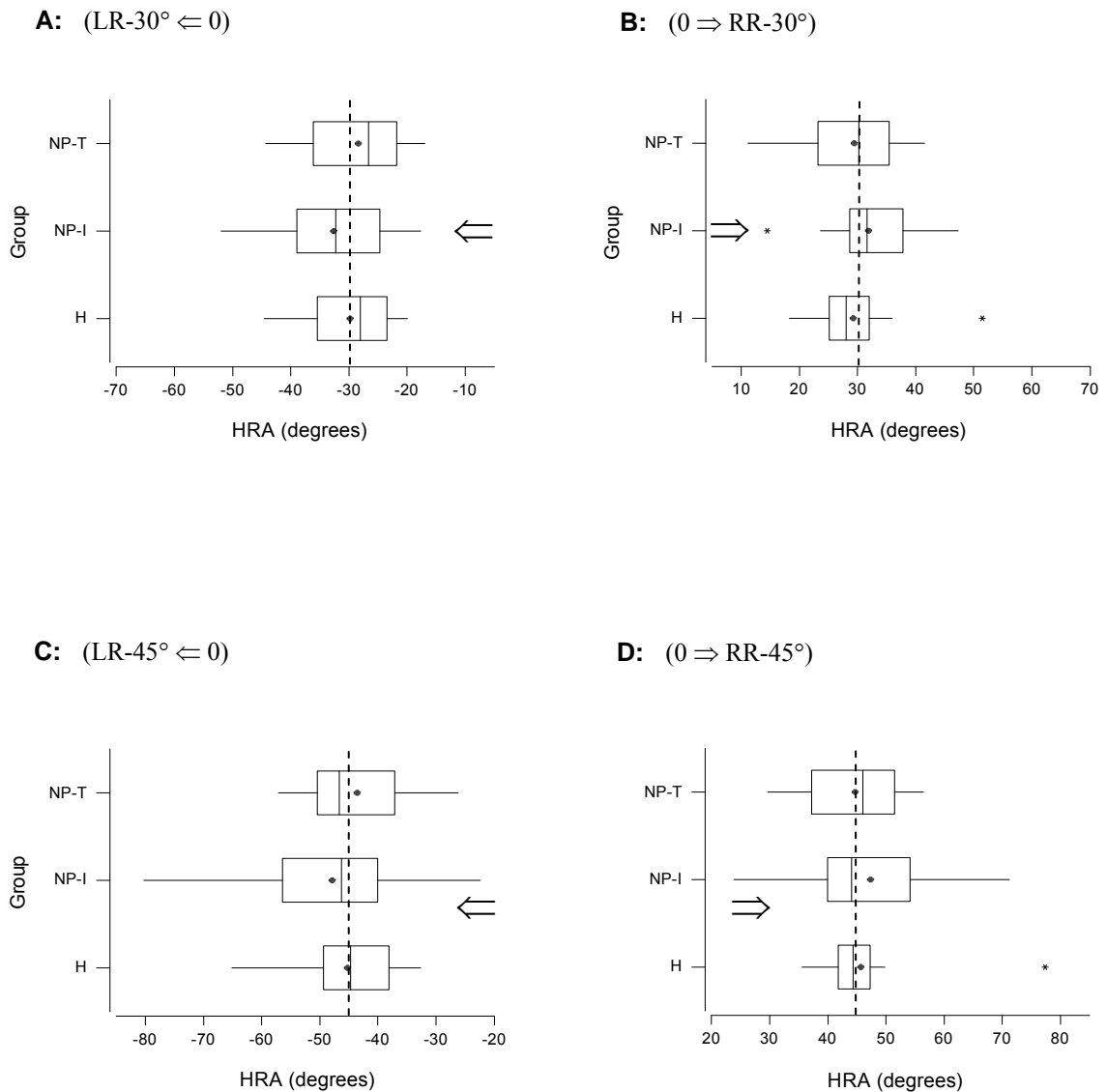
Abbreviations: LR  $\Rightarrow$  30, RR  $\Rightarrow$  30; LR  $\Rightarrow$  45, RR  $\Rightarrow$  45; positioning to 30° left & right and 45° left & right rotation:  $\theta Y-0$  represents the error compared with the NSA target angle (e.g. 30° or 45°) calibrated as 0°.

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA analysis. A statistically significant result is represented by  $P < 0.05$

\*\*Indicates F-ratio & P-value with Welch Correction.

**Head Movement Velocity and Range of Motion during kinaesthetic sensibility testing.**

The mean head movement velocities (both away and towards the repositioning target) for the NP-T and NP-I groups were very slightly lower in all trials compared with the control group (Table 64). No statistically significant difference was demonstrated for any of the movement characteristics using a one-way independent ANOVA.



**Figure 60.** Box plots of the signed repositioning error representing over/undershoot characteristics for the neck pain patients (insidious onset & traumatic; NP-I & NP-T) and healthy subjects (H). The horizontal ( $\theta_Y$ ) and vertical ( $\theta_X$ ) repositioning error (degrees) are displayed for L & RR  $\Rightarrow$  30 (A-B) and L & RR  $\Rightarrow$  45 (C-D) respectively. The middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25<sup>th</sup> and 75<sup>th</sup> percentile); the mean value is represented by the dot.; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:Q1 - 1.5 (Q3 - Q1) Upper Limit:Q1 + 1.5 (Q3 - Q1)).



**Table 64.** Comparison of movement characteristics during HRA-NSA in insidious (n=16) and traumatic (n=18) onset neck pain patients and healthy subjects (n=17) in the primary motion direction

Primary Motion	Movement Parameter	Healthy Subjects	Neck Pain (Insidious)	Neck Pain (Trauma)	df	F-test	P-Value*
		Mean ± SD	Mean ± SD	Mean ± SD			
LR-30	Velocity 1 (°/s)	-25.8 ± 7.65	-22.7 ± 5.91	-22.3 ± 6.65	2,48	1.40	0.26
	Velocity 2 (°/s)	27.6 ± 9.41	23.9 ± 5.87	22.2 ± 6.37	2,48	2.39	0.10
	Peak ROM (degrees)	-30.2 ± 7.52	-32.6 ± 9.14	-28.6 ± 8.18	2,48	1.02	0.37
LR-45	Velocity 1 (°/s)	-31.5 ± 11.61	-27.8 ± 6.76	-27.4 ± 9.42	2,48	0.98	0.38
	Velocity 2 (°/s)	35.4 ± 12.00	28.7 ± 6.07	28.8 ± 9.42	2,48	2.77	0.07
	Peak ROM (degrees)	-45.2 ± 9.09	-48.7 ± 12.78	-43.6 ± 8.83	2,48	1.05	0.36
-----							
RR-30	Velocity 1 (°/s)	25.3 ± 8.19	22.7 ± 5.91	22.3 ± 6.25	2,48	0.93	0.40
	Velocity 2 (°/s)	-26.9 ± 8.27	-24.2 ± 5.63	-23.1 ± 7.87	2,48	1.21	0.31
	Peak ROM (degrees)	29.4 ± 7.28	32.2 ± 7.55	29.7 ± 8.18	2,48	0.66	0.52
RR-45	Velocity 1 (°/s)	32.4 ± 10.42	27.1 ± 7.11	28.6 ± 9.77	2,47	1.45	0.25
	Velocity 2 (°/s)	-35.5 ± 11.87	-30.7 ± 6.35	-30.0 ± 11.52	2,47	1.41	0.26
	Peak ROM (degrees)	45.8 ± 8.95	47.8 ± 11.30	44.9 ± 8.54	2,47	0.39	0.68

Abbreviations: left and right rotation (LR, RR), extension (Ext), forward flexion (Flex); range of motion (ROM)

\*P indicates the level of significance for comparisons between groups using independent one-way ANOVA. A statistically significant result is represented by P < 0.05

\*\*Indicates F-ratio & P-value with Welch Correction.

### 6.3.3 Summary of Kinaesthetic Sensibility Results.

#### Test 1 - Head repositioning accuracy to subject straight ahead (HRA-SSA)

- **Absolute HRA:** No significant difference between control and symptomatic groups. However a significant difference was seen between the NP-I & NP-T groups for RR.
- **Signed HRA (overshoot/undershoot):** A significant difference between the control and NP-T for the Ext ⇒ 0 trial only (overshoot for symptomatic groups).
- **Peak ROM:** Reduced ROM for the NP-T group compared with healthy subjects but significant difference only for the Ext ⇒ 0 trial.
- **Velocity (mean):** Reduced head movement velocities away and toward target position in all trials for both neck pain groups. A significant difference was seen only with the Ext ⇒ 0 trial for the NP-T group compared with healthy subjects.

## **Test 2 - Head repositioning accuracy to subjective straight ahead with feedback**

### **(HRA-SSA[F])**

- **Absolute HRA:** No significant difference between control, NP-I and NP-T groups. For all groups, HRA was greater than with Test 1.
- **Signed HRA (overshoot/undershoot):** No significant difference between control, NP-I and NP-T groups.
- **Peak ROM:** Slightly reduced ROM for the NP-T group compared with healthy subjects but no significant difference.
- **Velocity (mean):** Reduced head movement velocities away and toward target position in all trials for both neck pain groups but no significant difference.

### **Test 3 - Head repositioning accuracy to non-straight ahead (HRA-NSA).**

- **Absolute HRA:** Slightly greater in in the plane of primary movement ( $\theta Y$ ) for the patient groups compared with the control group but no significant difference.
- **Signed HRA (overshoot/undershoot):** Tendency to overshoot target with NP-I group but no significant difference between control, NP-I and NP-T groups.
- **Velocity (mean):** Reduced head movement velocities away and toward target position in all trials for both neck pain groups but no significant difference.

## 6.4 DISCUSSION

This study focused on HRA performance in patients with chronic insidious onset neck pain and also patients with neck pain resulting from a ‘whiplash’ injury. This study also differed from previous studies in that the Zebris CMS 70P system was utilised for measurement of HRA. The overall results from the sample of patients in this study did not reveal any difference in HRA between the two subgroups of neck pain patients and the healthy age and sex matched controls with any of the three testing methods. These results are consistent with the earlier study in this thesis (insidious onset neck pain subjects) but conflict with previous studies where impairment in NHP repositioning has been found using various error parameters and measurement methods in patients with insidious neck pain (Kristjansson et al., 2003), ‘whiplash’ injury (Heikkila and Astrom, 1996; Heikkila and Wenngren, 1998; Kristjansson et al., 2003; Treleaven et al., 2003; Feipel et al., 2006; Treleaven et al., 2008), and groups where the neck pain onset has not been stated (Revel et al., 1991), or mixed (Humphreys and Irgens, 2002).

With respect to cervicocephalic kinaesthetic sensibility measurement utilizing repositioning and positioning tasks to non-neutral set points within the cervical range of motion, the results from the current study are in more agreement with other investigations (Table 2 – Chapter 1). Only one other study has looked at insidious onset patients with current pain and failed to find any meaningful differences compared with control subjects (Kristjansson et al., 2003). However, of the four that have included ‘whiplash’ subjects (Table 2 – Chapter 1), only two have demonstrated any significant impairment compared with control subjects (Loudon et al., 1997; Feipel et al., 2006).

### **Patient Groups and HRA**

Many of the possible reasons for these conflicting results are similar to those discussed in the earlier study in this thesis (Chapter 2). As also mentioned a focused comparison and discussion of the differences between results of these KS studies is often confounded by differences in patient groups, instrumentation, specific kinaesthetic tasks, measurement protocols and variations in the way the repositioning error is quantified. As such, only tentative overall conclusions can be drawn. Only two other studies have provided a comparison of patients currently experiencing neck pain of insidious onset with a healthy

control group (Kristjansson et al., 2003; Sjolander et al., 2008). With respect to patient group characteristics (neck pain severity, mean duration, gender etc.) there are several differences between the current results, those from the previous study in this thesis (Chapter 2), Kristjansson et al., (2003) and Sjolander et al., (2008) many of which appear contradictory with respect to the differences observed. The effect of these variables on KS has yet to be elucidated and requires further investigation.

### **Measurement Instruments, Protocols & HRA**

In finding no difference between patients in the earlier preliminary study (Chapter 2), one possible reason cited was differences in the measurement instruments and protocols between studies. Between the current and the earlier studies, several differences in testing methods existed; instrumentation (Zebris CMS 70P vs laser pointer), SSA target not found after a perpendicular active movement of the head, five repetitions, precoded instructions from speakers behind subjects. Although the populations studied were different, these changes did not appear to affect the discriminant value of the testing. When comparing between the current study and Kristjansson et al., (2003) and Sjolander et al., (2008), both studies measured 3-D error parameters remotely but utilised different devices; the current study using the Zebris CMS 70P and Kristjansson et al., (2003) and Sjolander et al., (2008), the 3-Space Fastrak system. The method agreement for measuring KS using these devices has not been established.

In the current study, the repositioning task to SSA ‘with repositioning’ between repetitions most closely resembled that used by Kristjansson et al., (2003) where a difference was found between patients and controls. However several possibly important differences still exist between the studies. The two NHP protocols used in this study were similar with that utilised earlier (Chapter 2) in that the repositioning target was defined as a subjective straight ahead position with respect to the trunk although the target position was not derived after a movement of the head in the perpendicular plane. This modified target position still differed to Kristjansson et al., (2003) who utilised the more commonly adopted NHP (Table 1 – Chapter 1). The subjects in the current study were also asked to perform a near-maximal amplitude head movement compared with a ‘full active rotation of the head within comfortable limits’ (Kristjansson et al., 2003). In addition, five repetitions were used arrive at the mean score compared with three used by Kristjansson et al., (2003). It is interesting to note that within Kristjansson’s study (2003) two completely different tests of assessing

NHP (one repositioning and one dynamic) failed to demonstrate differences. As highlighted in Chapter 2, it is still not clear what impact these variations have on the ability of the tests to detect abnormalities although in other patient populations ('whiplash' and mixed), SSA or NHP tests with differences in protocol similar to those highlighted above have been found to have some discriminative value (Table 1 – Chapter 1).

Aside from specific testing protocol differences between studies, the failure to demonstrate any meaningful difference between the insidious onset chronic neck pain to healthy control subjects in earlier study in this thesis (Chapter 2) also raised the possibility that testing methods other than SSA or NHP may be more of a proprioceptive challenge and have a more discriminative value. It is also theoretically possible that positions of the head-on-trunk within a range of motion may rely more on the cervical proprioceptive apparatus than vestibular information and a different nature of signalling which could reveal deficits in patients with mechanical cervical spine complaints. Loudon et al., (1997) demonstrated significant differences between 'whiplash' patients and control subjects as evidenced by a larger absolute error and a greater tendency to overshoot the targets in the patient group (Table 2 – Chapter 1). Kristjansson et al., (2003) used a similar method to Loudon et al., (1997), where the head was passively positioned to the horizontal target angle (30°) and the subject asked to actively return to this from the NHP whereas in the current study, the target angles were actively determined by the subjects (without feedback on accuracy) and then actively repositioned to this from a SSA position. In this sense, this test essentially assessed how well the subjects could find a target angle and then relocate to it. Kristjansson et al., (2003) also used a testing method derived from that used in the SPNTT (Gimse, Tjell, Bjorgen and Saunte, 1996) where the task is to relocate the trunk to an angle within an arc of motion. Armstrong et al., (2005) used a similar active protocol and also failed to demonstrate a difference between patient and controls. Fiepel et al., (2006) did demonstrate a significantly larger error for 'whiplash' subjects compared with controls but the actively derived nSA target was guided by feedback from the investigator. In finding no impairment for the insidious onset neck pain patients compared with controls, the current results and those of Kristjansson et al., (2003) superficially suggest that despite producing larger repositioning errors compared with SSA or NHP tasks, all these testing methods may lack a useful discriminative value for this neck pain subpopulation. It is interesting to note that larger rotation target angles (45° vs 30°) did not provide any further discriminative value compared with Kristjansson et al., (2003). Loudon et al., (1997) utilised 30° and 50° angles

but it was not stated if there were any differences between these methodological approaches.

The results from the current study may not be entirely supportive of the position that these non-NHP tests lack discriminative value as contrary to Kristjansson et al., (2003), no impairment was also found with this insidious onset neck pain group using a NHP technique. This may suggest that the sample population in this study lacked any deficits in KS. As such, it is possible that the repositioning test to a non-neutral position used in this study may have some discriminative value in another subgroup of patients with insidious onset pain. However, the fact that Kristjansson et al., (2003) also failed to find any impairment with albeit significantly different NHP tests, may reinforce the position that differences in testing methods may be the source for conflicting results. At this stage, no firm conclusions can be drawn without further study of the variables concerned.

#### **‘Whiplash’ Patients & HRA**

The results for the ‘whiplash’ subjects are also quite contradictory. If the NHP tasks are again considered first, only two of the several investigations of this subgroup of chronic neck pain patients (Table 1 – Chapter 1) have failed to reveal any impairment using these tests compared with asymptomatic control subjects (Kristjansson et al., 2003; Armstrong et al., 2005). As highlighted earlier, differences in patient characteristics is a potential reason for these contradictions. Many of these have categorised the severity of the ‘whiplash’ injury using the WAD criteria. The effects of WAD severity on HRA has yet to be clarified. However, based on this classification, most of those used by Armstrong et al., (2005) could be graded as WAD II and thus not very different to those where abnormality has been reported.

A closer evaluation of the questionnaires completed by patients in the current study at time of KS measurement revealed that a few subjects could not clearly relate the onset of their problem to ‘whiplash-type’ injury. They had all experienced a ‘whiplash’ event in the past but it was not clear whether it occurred at onset or was a contributing factor. This ambiguity is a methodology flaw and certainly highlighted that information from patient questionnaires and case notes may differ. If the group were considered mixed in onset (or insidious), the results are still contrary to other studies (Revel et al., 1991; Humphreys and Irgens, 2002; Kristjansson et al., 2003).

As with the insidious onset neck pain subjects, measurement instrumentation, specific kinaesthetic tasks, measurement protocols are again also possible factors responsible for the contradictory results for the ‘whiplash’ patients. The results from Kristjansson et al., (2003) could provide some evidence that differences in testing methodology may be a factor for the somewhat contradictory results between studies. Impairment in KS involving a NHP target was only demonstrated using one of the three testing methods. These involved a more complex repositioning task to a NHP after fig-8 shaped movement of the head and a dynamic test of relocation accuracy through the starting NHP position. However, the tests that failed to show any differences between patients and control subjects were markedly different in nature to those utilised in the current study and by Armstrong et al., (2005). Although there were differences in task, protocol and testing instrumentation, the only test that revealed abnormalities used by Kristjansson et al., (2003) more closely resembled that used in the current study, the more widely used and ‘traditional’ test of relocation to NHP or SSA.

The results for the non-straight ahead tasks are in agreement with Kristjansson et al., (2003) and Armstrong et al., (2005) but conflict with two other investigations (Loudon et al., 1997; Feipel et al., 2006). This could suggest that non-SSA tasks may not have the same discriminative value for patients suffering persistent pain from a ‘whiplash’ injury compared with the more traditional and widely utilised NHP tests. Once again, patient characteristics and testing methods are the factors most likely responsible for the contradictory results although both Kristjansson et al., (2003) and Loudon et al., (1997) utilised a very similar testing protocol in their respective studies. It is interesting to note that the patient group studied by Loudon et al., (1997) had experienced up to three ‘whiplash’ injuries in the preceding two years; possibly a greater injury profile compared than two other studies and the current investigation. However, larger errors in repositioning were demonstrated in a ‘whiplash’ group (Feipel et al., 2006) with a similar clinical profile to the groups that failed to reveal any deficits (Kristjansson et al., 2003; Armstrong et al., 2005).

### **Head Movement Velocity and HRA**

In finding no difference between patient and control subjects, speed of head movement and range of motion during the tests were highlighted in Chapter 2 as factors that may affect the the discriminant value of the tests between groups and indeed whether the tests are utilizing proprioceptive information from the cervical spine. Concurrent measurement of these

variables during KS tasks have not been measured or reported in previous studies although one investigation did control the speed of head movement when positioning the head to a non-NHP target (Teng et al., 2007). For HRA-SSA, the results from the current study demonstrated a trend for the patient groups to move the head slower both away and towards the repositioning target for both SSA tasks compared with the control subjects. For test 1, a significantly reduced head movement velocity (outward and return) was noted between the 'whiplash' and healthy control group for the Ext  $\Rightarrow$  0 task. It is also interesting to note that the the speed of movement was generally slower for both patient groups for the vertical plane (Flex/Ext) repositioning tasks compared with the horizontal movement tasks. The lack of a more general statistically significant difference between groups may be related to the relatively small sample size and a large variance.

For the head-on-trunk repositioning tasks within the horizontal range of motion (test 3), it is interesting to note that the control and both neck pain groups moved slower than with the horizontal plane HRA-SSA tasks (Table 64). The results from the current study also demonstrated a trend for the patient groups to move the head slower both away and towards the repositioning target for all tasks compared with the control subjects with the nSA tasks. However, these differences were not statistically significant. In addition the results revealed that the subjects in all groups tended to move faster with the 45° repositioning tasks compared with the 30° tasks and that they tended to move slower toward the non-neutral target positions than when returning to the SSA position. As discussed in Chapter 2, it is theoretically possible that if the two groups in a study generally move and relocate at different speeds, the results from a slower moving group may reflect the function of the cervical mechanoreceptive apparatus as opposed to the vestibular system with faster movements (Mergner et al., 1983). This could potentially affect the error sizes and patterns and difference between groups. However, when patients have been asked to move the head as fast as possible, deficits have been demonstrated in patient groups (Sjolander et al., 2008). The speed of head movement velocity that best reflects the function of cervical mechanoreceptive apparatus has not been identified. Two studies have utilised a speed of movement slower than 35°/sec (Lee et al., 2007; Teng et al., 2007). Teng et al., (2007) derived this head movement speed from a study investigating age related modulation of vestibulo-ocular reflex (Goebel, Hanson and Fishel, 1994) not cervical proprioceptive function. The study also failed to demonstrate any HRA in their neck pain population (Teng et al, 2007).



## **Cervical Spine Range of Motion and HRA**

The results from this study also revealed a slightly reduced range of motion for the ‘whiplash’ group in both tests 1 and 2 (tests of HRA-SSA) although a statistical significance was again found for the Ext  $\Rightarrow$  0 task with test 1. The impact of these variables on HRA and the discriminative value of the tests is currently unclear with some conflicting results (Sjolander et al., 2008). However it interesting to note with the current study, for the Ext  $\Rightarrow$  0 task (test 1), a statistically significantly greater overshoot was noted for the ‘whiplash’ group compared with the insidious onset neck pain group and healthy control subjects (Table 57). An inspection of the boxplots, scatterplots and CI’s for the differences suggests that the clinical significance of this is questionable. These results also suggest that the potential sensitization effect of a larger range of motion during the testing with resultant overshooting may not have been a factor with this study (Lee et al., 2005). Further research is needed to evaluate the effect of ROM and speed of head movement on various KS tasks, in particular, the discriminative value of the tests and whether these variables affect the ability to detect abnormalities in the neck mechanoreceptors and related pathways.

## **Control Subjects and HRA**

As discussed in Chapter 2, another consideration when examining the lack of difference between patient and control subjects in the present study is a poor HRA performance from the control subjects i.e., they don’t reflect a true normal population. There were no clear outlying performances from this group compared with the patient groups which could significantly affect the relatively small sample size in this study. One potential issue that may lead to variability of HRA both within and between studies is subject motivation. Both patients and control subjects may vary the amount of effort and attention that they put into achieving the target positions based on subjective factors. In these studies we took every effort to make the experience as pleasant and easy as possible and received a high level of patient compliance. As highlighted earlier, a comparison with normative performance from other studies is difficult due to differences in methodologies and measurement equipment and set-up. It is also sometimes hard to compare in detail as many studies have not provided all the measurement parameters used in this study or presented them in different units (Table 1 – Chapter 1). Despite differences in measuring devices and measuring protocols, for healthy subjects using remote measuring devices, the results from the present study are similar to others for SSA or NHP repositioning tasks (horizontal rotation) (Table 65) and

for non-neutral to repositioning tasks to 30° horizontal rotation (Table 66). It is interesting to note that for the non-straight ahead tasks, the variance for the healthy subjects was larger in the present study than with the other studies. This may reflect that in the current study the target position was determined by the subjects without guidance compared with a head being passively positioned by the examiner in the other studies.

**Table 65.** A comparison of the present vs other studies HRA-SSA/NHP for healthy control subjects. Data represents absolute  $\theta Y$  error for horizontal rotation repositioning tasks (Mean  $\pm$  SD)

	Present Study Test 1	Present study Test 2	Earlier HRA study (Chapter 2)	Kristjansson, 2001	Kristjansson, 2003	Sterling, 2003	Treleavan, 2003	Armstrong, 2005
LR $\Rightarrow$ 0	2.3 $\pm$ 1.3	1.2 $\pm$ 0.8	2.5 $\pm$ 2.6	2.5 $\pm$ 1.3	2.5 $\pm$ 1.1	2.6 $\pm$ 0.3*	2.0 $\pm$ 0.3*	3.3 $\pm$ 2.32
RR $\Rightarrow$ 0	2.4 $\pm$ 1.6	1.3 $\pm$ 1.3	5.2 $\pm$ 5.1			2.7 $\pm$ 0.3*	2.5 $\pm$ 0.2*	

*Abbreviations:* left and right rotation (LR, RR),  
\*SEM

**Table 66.** A comparison of the present vs other studies HRA to non-straight ahead positions for healthy control subjects. Data represents absolute  $\theta Y$  error for horizontal rotation repositioning tasks to 30° (Mean  $\pm$  SD)

	Present Study	Loudon, 1997	Kristjansson, 2001	Kristjansson, 2003
LR $\Rightarrow$ 30°	6.1 $\pm$ 4.3	5.0 $\pm$ 1.8	5.9 $\pm$ 3.8	5.5 $\pm$ 2.0
RR $\Rightarrow$ 30	5.1 $\pm$ 5.2			

*Abbreviations:* LR  $\Rightarrow$  30, RR  $\Rightarrow$  30; positioning to 30°

### Comparison between HRA tests

Although all of the three tests of HRA used in the current study failed to demonstrate a difference between patients and healthy control subjects, a comparison between the HRA results obtained with each test provides some useful observations. The absolute errors obtained with the tests to non-straight ahead positions were clearly larger than for the two tests of HRA-SSA. This pattern of absolute error size is consistent with other studies that have directly compared HRA performance using these tests (Kristjansson et al., 2001; Kristjansson et al., 2003; Feipel et al., 2006). These results suggest that the non-straight ahead positions are perhaps more challenging despite being no more discriminative. The

results from this study also demonstrated that the error obtained when the head is repositioned to the target after each repetition is smaller than without this testing procedure (Table 65) although again no difference in discriminative value was seen with the current study.

## 6.5 CONCLUSION

The results of this study suggest that patients with chronic neck pain of both insidious onset (non-traumatic) and from a 'whiplash' injury show little evidence of impaired cervicocephalic kinaesthetic sensibility, when measured using subjective straight-ahead and non-straight-ahead HRA tests. Although for HRA-SSA, these results for the insidious onset neck pain patient group are in agreement with the earlier study in this thesis, the results for both patient groups using this test method were generally not consistent with the majority of previous published studies. For the nSA tests, the results of the current study are in slightly closer agreement with previous studies looking at differences between patients and control subjects although the result from these studies are mixed and quite contradictory.

The conflicting results between the studies may reflect a difference in the patient populations highlighting further the possible heterogeneity of chronic mechanical neck pain. It has been tentatively suggested that the pattern of sensorimotor disturbances in chronic neck is individual, and not always related to the aetiology of the pain (Sjolander et al., 2008). The somewhat contradictory results may also reflect differences in methodology and specific aspects of measurement protocol and sample size. In further studies, it would be useful to look at the effects of neck pain patient characteristics (e.g., pain severity, chronicity, disability & aetiology) on cervicocephalic KS measured using tests of HRA-SSA and non-straight ahead positions. There is also a need to investigate the effects of movement speed and range of motion during the testing procedure on both error size and discriminative value of the tests.

This study highlights the need to compare further the effects of differing measurement protocols on kinaesthetic performance and the relative discriminative value of the tests. This will help identify a solid test or tests which can consistently detect deficits in cervicocephalic kinaesthetic sensibility in patients with persistent painful mechanical disorders of the cervical spine.

## Chapter 7

---

### GENERAL DISCUSSION

---

Despite a number of studies demonstrating deficits in cervicocephalic kinaesthetic sensibility in patients with chronic ‘mechanical’ neck pain (Table 1-3; Chapter 1), an appraisal of the current evidence shows that there is still no solidly established best measurement method for the routine evaluation of mechanoreceptive somatosensory (proprioceptive) function using head on trunk positioning and movement tasks. The results from the studies in this thesis have added to this conflicting data and served to highlight some of the possible problems in this field of research and clinical practice. Variations in testing protocols including the possibility of different levels of subject motivation, and differences in measurement equipment and group characteristics are the main reasons for the difficulties in collating data and making recommendations. Differences in error variables and incomplete description of the group characteristics have added to this problem.

The validity and discriminative value of the more traditional tests of HRA (Tables 1 & 2) have been questioned by several authors suggesting alternative approaches to the assessment of cervical spine proprioception (Kristjansson et al., 2001; Kristjansson et al., 2003; Kristjansson et al., 2004; Lee et al., 2005). However, from the research to date repositioning tasks to a SSA/NHP appear to be the most discriminative between patient and control groups (Kristjansson et al., 2003). What is also clear is that the clinical characteristics of the various testing procedures have not been adequately assessed in particular issues of *reliability* and measurement agreement between the numerous protocols and measurement devices and methods that have appeared. Here we demonstrate that the multiple test *random error* effects on HRA are of a similar magnitude to variations between control and patient groups. Therefore, these tests can only have clinical utility where the patient HRA effect is considerably larger than observed in these studies. In addition, most studies including patient populations have involved chronic neck pain subjects. However, what isn’t clear is when deficits in cervicocephalic kinaesthetic sensibility appear and what interventions may be best for this impairment assuming that an impairment in KS is a primary pathological component requiring a direct treatment protocol.

### **Validity of HRA tests: Cervical vs Vestibular function**

Although the validity of the KS tests has not received much direct investigation, abnormalities with the traditional KS tests are thought to reflect altered cervical afferent input or an alteration in afferents' integration and tuning (Revel et al., 1991). A number of experimental arguments point toward a cervical spine role. Indeed, it is often assumed that when patients have both 'mechanical' neck pain and KS impairment, it is reasonable to assume a cervical cause for the deficit (Revel et al., 1991).

The abundance of cervical mechanoreceptors and their central and reflex connections reflects the important proprioceptive role of the cervical spine (Treleaven et al., 2008). The peripheral receptors can be grouped according to anatomical location; joint, muscle, and cutaneous (Armstrong et al., 2008). The muscle mechanoreceptors often take the form of complex arrays of muscle spindles, golgi tendon organs, paciniform corpuscles and free nerve endings (Richmond, 1988).

In human cervical facet joint capsules, small numbers of mostly type II mechanoreceptors have been found (McLain, 1994). In contrast to the muscle spindle receptors, little difference in the distribution of the receptors within the upper and lower cervical spine facet capsules was observed. The afferent information from these receptors might have a localised effect in feeding back to the small deep muscles of the neck providing stability for the motion units. They may also provide a complimentary role to muscle receptors in signalling position sense at extremes of movement when a joint is distracted or compressed (Armstrong, McNair and Taylor, 2008).

It is commonly felt that in signalling muscle length and change of length, the muscle spindle is primarily responsible for signalling the cervical proprioceptive information involved in the conscious perception of equilibrium, position and spatial orientation the sensory function/s evaluated with the cervicocephalic KS tests (McCloskey, 1978; Armstrong et al., 2008). The joint and cutaneous receptors are likely to provide supplementary information to the CNS. In humans, high densities of spindles have been found in many of the cervical muscles (Amonoo-Kuofi, 1983; Boyd-Clark, Briggs and Galea, 2002; Liu, Thornell and Pedrosa-Domellof, 2003). A particularly high density of receptors have been observed in the deep muscles forming the suboccipital triangle (the

superior and inferior oblique capitus, and rectus capitus posterior major and minor (Peck, Buxton and Nitz, 1984; Kulkarni, Chandy and Babu, 2001).

The central connections and processing of the cervical mechanoreceptive input is complex (Armstrong et al., 2008; Treleaven et al., 2008). It is thought that discharge patterns from muscle spindles are relayed to the CNS and processed in tandem with vestibular input. If positional information from either source is inaccurate or fails to be appropriately integrated in the CNS, error in head positioning may occur. The cerebellum and cerebral cortex also have a role to play providing feed forward and modulatory influences on various head movements depending on the task.

It has been suggested that the cervical spine is the primary source of afferent information for HRA tasks as the threshold for head positioning does not differ significantly if the head moves relative to the trunk or the trunk relative to the head (Taylor and McCloskey, 1988; Revel et al., 1991). In addition when subjects were asked to grade their perception of various movements, the results suggest that the head movements may utilise peripheral kinaesthetic information primarily from the cervical spine mechanoreceptive apparatus (Mergner et al., 1983). Studies involving posterior neck muscle vibration as a cervical proprioceptive stimulus have also demonstrated an influence of cervical proprioception on perceived subjective visual straight ahead (Lackner and Levine, 1979; Biguer, Donaldson, Hein and Jeannerod, 1988; Taylor and McCloskey, 1988; Karnath, Sievering and Fetter, 1994; Strupp, Arbusow, Borges Pereira, Dieterich and Brandt, 1999) and head and trunk midline position (Ceyte, Cian, Nougier, Olivier and Roux, 2006). However, although supportive, these results do not provide direct quantification of the cervical contribution to head position and movement with the specific tasks utilised in the various cervicocephalic kinaesthetic sensibility studies.

As highlighted above, the validity of the tests of cervicocephalic kinaesthetic sensibility for measuring cervical mechanoreceptive afferent function (and/or central integration) has not been directly established. As introduced in Chapter 1, with two exceptions (Kristjansson et al., 2001; Kristjansson et al., 2003) nearly all the investigations attempting to quantify cervical mechanoreceptive function have utilised only kinaesthetic performance tests involving movements of the head on a stationary trunk. Assuming that conscious, purposeful head movement kinaesthetic tasks (with vision occluded) are reliant on

peripheral afferent information, the procedure of active head-repositioning to a specific position with respect to the trunk potentially involves ‘head-in-space’ information from the vestibular system and ‘head-on-trunk’ proprioceptive information from the cervical spine mechanoreceptors (Mergner et al., 1983; Mergner et al., 1991; Karnath et al., 1994). As such, the performance of these tests may not represent a specific measure of cervical spine afferent function.

The afferent role of the vestibular apparatus with the various forms of head-on-trunk cerviccephalic KS tests also remains unclear and until recently had not been studied directly (Treleaven et al., 2008). The ‘head-in-space’ vestibular afferent information that may potentially influence the tests could arise from the semicircular canal or otolith systems within the vestibular apparatus (Kandel, Schwartz and Jessell, 2000). These systems could be affected in neck pain patients who have experienced direct or indirect head trauma (Chester, 1991; Kogler, Lindfors, Odkvist and Ledin, 2000; Basta, Todt, Scherer, Clarke and Ernst, 2005; Ernst, et al., 2005). There are five vestibular end organs within each ear; three semicircular canals that transduce angular acceleration and two otolith organs (utricle and saccule) that transduce linear acceleration and head tilt with respect to gravity. From a rather simplistic viewpoint, head positioning tasks to a static straight-ahead target or an angle within and arc of head motion would most theoretically utilise positional gravity referenced afferent signals from the otolith system. The orientation of the utricle and saccule would suggest that any contribution to the perception of NHP or SSA would occur primarily in the sagittal (vertical or flexion-extension) and coronal (latero-flexion) planes. As such it has been suggested that KS tasks in the transverse (horizontal rotation) plane would minimise the input from these vestibular organs and more specifically isolate cervical proprioceptive information (Revel et al., 1994).

Although the semicircular canals are primarily involved with angular acceleration, a contribution of these organs to the KS tasks, particularly dynamic ones, cannot be excluded (Hamann, Strauss, Kellner and Weiss, 1992; Kristjansson et al., 2004). There is convergence of otolith and semicircular canal input at all central vestibular levels, from the vestibular nuclei to the vestibular cortex (Kingma, 2006). Indeed, recent reports indicate that canal information is used together with the otolith input in an internal model of the brain (Green, Shaikh and Angelaki, 2005; Merfeld, Park, Gianna-Poulin, Black and Wood, 2005). As the utricle and macule are unable to discriminate translations and tilts, functional



synergy between the two vestibular systems may play a role with some perceptual tasks such as dynamic head tilt (Kingma, 2006). In addition, caloric stimulation of the horizontal semicircular canals can alter the perception of perceived subjective visual straight ahead (Karnath et al., 1994).

It is thought that HRA tasks that involve slow head movements may be below threshold for the vestibular apparatus are functionally best suited to the cervical proprioceptive apparatus (Lark and McCarthy, 2007). With respect to the semicircular canals, slow head movements may not be able to overcome the inertia of the cupula in the semicircular canals (Mergner et al., 1983; Kristjansson et al., 2004). As yet, it is not clear if head movement speed determines which afferent information is preferentially utilised and if it is a factor, what speed of head movement would be best for isolating the cervical mechanoreceptive apparatus.

A useful model for investigating the role of the vestibular system in cervicocephalic sensibility tasks, may be to study HRA in subjects with primary vestibular system pathologies with no neck complaints (Kingma, 2006; Wuyts, Furman, Vanspauwen and Van De Heyning, 2007) or the effect of cervical and vestibular stimulations or challenges during the testing procedures. To the author's knowledge, this has only recently been done. A comparative study between subjects with 'whiplash' and those with an ongoing unilateral vestibular pathology (acoustic neuroma) revealed greater head repositioning errors for both groups compared with control subjects but no difference between the two patient groups (Treleaven et al., 2008). These results could suggest that HRA-SSA testing may be useful in determining sensorimotor control abnormalities due to mismatched afferent input from one of either abnormal vestibular and/or cervical origins. However, the testing may not be useful to differentiate between a vestibular and cervical cause of the disturbances. This implies that the tests may not be a specific measure of cervical afferent function as previously thought. Several study limitations may exist though including an inadequate sample size, concomitant neck involvement in subjects with acoustic neuroma, or concomitant vestibular abnormality in patients with a 'whiplash' injury (Treleaven et al., 2008).

It is clear that direct investigations aimed at identifying the role of the all the potential proprioceptive and vestibular afferent components in head repositioning tasks are needed

before more solid conclusions can be reached. Specifically, it needs to be established whether modification of the HRA-SSA test could assist in differentiating a cervical versus a vestibular component. This line of investigation would be particularly helpful in designing methodologies to best identify a cervical proprioceptive abnormality and in identifying the discriminant value of cervicocephalic kinaesthetic sensibility tests for patients with neck pain. This would be especially important when the neck pain followed a 'whiplash' injury where concomitant otolith system dysfunction could be present (Chester, 1991; Kogler et al., 2000; Ernst et al., 2005). This research could also help inform the usefulness of cervicocephalic KS tasks in determining the nature of pathology in patients with neck pain and dizziness/dysequilibrium (Treleaven et al., 2006). Deficits in HRA-SSA/NHP have been demonstrated in patients with dizziness and dysequilibrium of a suspected somatosensory cervical spine origin (Heikkila et al., 2000; Treleaven et al., 2003; Treleaven et al., 2006). Where these symptoms exist, HRA may be worse than for patients with simple neck pain (Heikkila and Wenngren, 1998). However, somato-sensory cervical dizziness (SSCD) is often a diagnosis of exclusion and as such, investigation of vestibular function with these patients in particular may go some way to clarifying the validity of the KS tests.

Where there is no evidence of a concurrent head injury, primary injury to the vestibular apparatus (insidious onset neck pain subjects) or associated dizziness or dysequilibrium, investigators and clinicians are generally confident to assume that impairments in KS when seen in conjunction with mechanical neck pain, most likely reflect abnormal cervical afferent input to the postural control system or central integration (Revel et al., 1991; Treleaven et al., 2006). In addition to this and the experimental evidence highlighted above, improvements in kinaesthetic performance with various direct, passive cervical spine treatment regimes go some way to reinforcing the construct and content validity of the testing procedure and as a measure of cervical afferent information and/or integration (Rogers, 1997; Heikkila et al., 2000; Palmgren et al., 2006).

### **Pain & HRA/Proprioception**

Although HRA has been shown to improve in the absence of a reduction in pain perception (Heikkila et al., 2000), the direct effect of nociception and CNS processing of nociceptive information on cervicocephalic KS cannot be ignored.

Pain intensity has been the major focus for investigations into the relationship between pain and HRA. Studying the intensity of pain can be quite difficult especially if it occurs intermittently or subclinically as may be the case in the patient based studies in this thesis. The study in Chapter 2 of this thesis did not show any relationship between pain intensity and HRA although the subject numbers were low in this study. Other studies have also failed to consistently show any relationship between pain intensity and cervicocephalic sensibility tests (Revel et al., 1991; Heikkila and Astrom, 1996; Kristjansson et al., 2003; Armstrong et al., 2005; Lee et al., 2007). However, a recent study has demonstrated an association between pain frequency rather than intensity in individuals with subclinical neck pain (Lee et al., 2007).

It is clear from animal studies at the lumbosacral levels that nociception is quite capable of inducing changes in muscle spindle discharge and the proprioceptive properties of brainstem neurons (Johansson, Djupsjobacka and Sjolander, 1993; Thunberg, Ljubisavljevic, Djupsjobacka and Johansson, 2002). In a cervical spine animal model, excitation of chemosensitive nociceptors in facet joints and muscles induces reflex activation of fusimotor neurones which alter the static and dynamic sensitivity of the muscle spindles (Thunberg, et al., 2001). If the signal-to-noise ratio of the spindles afferents is reduced in response to increased noise in the fusimotor signal, it would be reflected in decreased proprioceptive acuity (Sjolander et al., 2008). It could be argued then that reduction of pain alone may reduce any proprioceptive abnormality on its own. However, it could also be argued that proprioceptive abnormality may lead to aberrant sensorimotor control of cervical spine movements exposing the spinal components to abnormal and repetitive strain (O'Sullivan, 2005; Panjabi, 2006). Although this model is taken from the low back arena, if it holds true for the cervical spine, then it is quite possible that neck pain and proprioceptive deficiency may both sustain and perpetuate each other.

The presence of chronic nociception and possible subsequent CNS neuroplastic changes in areas of the neuraxis involved in the mental representation of cervical posture and spatial awareness theoretically may also add a further layer of complexity to the problem (Apkarian et al., 2004). This may have a descending modulation effect on voluntary sensorimotor control and the numerous cervical and vestibular reflex interactions.

## Chapter 8

---

### OVERALL CONCLUSIONS

---

Since Revel et al., (1991) originally introduced the head repositioning accuracy test, numerous kinaesthetic sensibility studies have been published (Chapter 1; Tables 1-3) looking at a variety of testing methods in various patient populations. Despite this diagnostic research and the clinical application of the results (e.g., intervention studies), a test that can be routinely applied in the clinical setting for the purposes of diagnosis and monitoring treatment effects has still not been established. The studies to date suggest that an active HRA test to SSA/NHP that is established by the patient may have the greatest discriminative value (Table 1) compared with nSA target tests (Table 2). Tests of more complex sensorimotor movement tasks (Table 3) that may involve more unfamiliar head control skills may prove useful but at present the evidence is limited (Kristjansson et al., 2004).

It is unlikely that the KS tests identify specific subpopulations of neck pain patients (e.g, insidious vs ‘whiplash’) as there have been several contradictory studies many of which have shown considerable overlap between the patient and healthy groups. It appears that patients with chronic neck pain who experience pain more frequently are more likely to exhibit a HRA deficit (Lee et al., 2007). It may also be that patients suffering chronic pain following a ‘whiplash’ injury may be more likely to exhibit HRA deficits compared with other patient populations. However, the data is far from conclusive at this stage and beyond this broad clinical profile, the more specific clinical profile of a patient who may exhibit HRA deficits is unclear.

It is also unlikely that HRA tests represent a unique test of cervical proprioceptive function (peripheral or central integration) as previously thought and therefore a test exclusive to neck disorders. Although the evidence is limited, it would appear that results of an HRA test utilising SSA/NHP target may be useful in determining sensorimotor control abnormalities due to the mismatched afferent input from one of either abnormal vestibular and/or cervical origins (Treleaven et al., 2008). As such, the HRA tests may not be useful to differentiate between a vestibular and cervical cause of any deficit. Therefore, the tests as

they stand are unlikely to represent a specific test for cervicogenic somatosensory dizziness. Added to this, it is also possible that HRA deficits seen with ‘whiplash’ patients who are not experiencing any dizziness or dysequilibrium, may be the result of an otherwise undetected or observed vestibular system deficit. From this, it is clear that further investigation is needed into the mechanisms involved in head repositioning errors and specifically whether a modification of the tests could assist with differentiating a cervical versus a vestibular component.

It is clear that the results from the tests as they stand should be interpreted carefully in the clinical environment when a vestibular abnormality could be present. Leading on from this, a standardised method or methods need to be established that have a cervical spine discriminative value. This may involve the HRA tasks commonly utilised or more complex non-learned movements. The sensitivity and specificity of the tests needs to be determined together with their predictive value. The *reliability* also needs to be further established using appropriate analysis methods. This will have particular value when looking at the ability of the tests to monitor changes in HRA over time.

As the populations studied appear to be heterogeneous, it is important, that a solid normative data base is established with the useful HRA tests. This database may help in assessing the performance of patient groups by providing a baseline data set of KS. This would be particularly useful for the clinician to establish any abnormality present in a particular patient and as an aid to determining the most specific and effective approach to treatment.

## References

1. Allison, G.T. and Fukushima, S. (2003). Estimating three-dimensional spinal repositioning error: The impact of range, posture, and number of trials. *Spine* 28 (22), 2510-2516.
2. Altman, D.G. and Bland, J.M. (1983). Measurement in medicine: The analysis of method comparison studies. *Statistician* 32, 303-317.
3. Amonoo-Kuofi, H.S. (1983). The density of muscle spindles in the medial, intermediate and lateral columns of human intrinsic postvertebral muscles. *J Anat* 136 (Pt 3), 509-519.
4. Anastasopoulos, D., Mergner, T., Becker, W. and Deecke, L. (1991). Sensitivity of external cuneate neurons to neck rotation in three-dimensional space. *Exp Brain Res* 85 (3), 565-576.
5. Andary, M.T., Hallgren, R.C., Greenman, P.E. and Rechten, J.J. (1998). Neurogenic atrophy of suboccipital muscles after a cervical injury: A case study. *Am J Phys Med Rehabil* 77 (6), 545-549.
6. Apkarian, A.V., Sosa, Y., Sonty, S., Levy, R.M., Harden, R.N., Parrish, T.B. and Gitelman, D.R. (2004). Chronic back pain is associated with decreased prefrontal and thalamic gray matter density. *J Neurosci* 24 (46), 10410-10415.
7. Aprill, C. and Bogduk, N. (1992). The prevalence of cervical zygapophyseal joint pain. A first approximation. *Spine* 17 (7), 744-747.
8. Armstrong, B., McNair, P. and Taylor, D. (2008). Head and neck position sense. *Sports Med* 38 (2), 101-117.
9. Armstrong, B.S., McNair, P.J. and Williams, M. (2005). Head and neck position sense in whiplash patients and healthy individuals and the effect of the cranio-cervical flexion action. *Clin Biomech (Bristol, Avon)* 20 (7), 675-684.
10. Atkinson, G. and Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 26 (4), 217-238.
11. Balague, F. (2008). The bone and joint decade (2000-2010) task force on neck pain and its associated disorders: A clinician's perspective. *Spine* 33 (4 Suppl), S4.
12. Barnsley, L., Lord, S.M., Wallis, B.J. and Bogduk, N. (1995). The prevalence of chronic cervical zygapophysial joint pain after whiplash. *Spine* 20 (1), 20-25; discussion 26.
13. Basta, D., Todt, I., Scherer, H., Clarke, A. and Ernst, A. (2005). Postural control in otolith disorders. *Hum Mov Sci* 24 (2), 268-279.
14. Bastian, H.C. (1888). The "Muscular sense"; its nature and cortical localisation. *Brain* 10, 1-137.
15. Berthelsen, P.G. and Nilsson, L.B. (2006). Researcher bias and generalization of results in bias and limits of agreement analyses: A commentary based on the review of 50 acta anaesthesiologica scandinavica papers using the altman-bland approach. *Acta Anaesthesiol Scand* 50 (9), 1111-1113.
16. Biguer, B., Donaldson, I.M., Hein, A. and Jeannerod, M. (1988). Neck muscle vibration modifies the representation of visual motion and direction in man. *Brain* 111 ( Pt 6), 1405-1424.
17. Bland, J.M. and Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1 (8476), 307-310.
18. Bland, J.M. and Altman, D.G. (1990). A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 20 (5), 337-340.
19. Bland, J.M. and Altman, D.G. (1996a). Measurement error. *Bmj* 313 (7059), 744.
20. Bland, J.M. and Altman, D.G. (1996b). Measurement error and correlation coefficients. *Bmj* 313 (7048), 41-42.
21. Bland, J.M. and Altman, D.G. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res* 8 (2), 135-160.
22. Bland, J.M. and Altman, D.G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound Obstet Gynecol* 22 (1), 85-93.
23. Boden, S.D., Mccowin, P.R., Davis, D.O., Dina, T.S., Mark, A.S. and Wiesel, S. (1990). Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation [see comments]. *J Bone Joint Surg [Am]* 72 (8), 1178-1184.
24. Bogduk, N. and Marsland, A. (1988). The cervical zygapophysial joints as a source of neck pain. *Spine* 13 (6), 610-617.
25. Bogduk, N. (1995). The anatomical basis for spinal pain syndromes. *J Manipulative Physiol Ther* 18 (9), 603-605.
26. Bolton, J.E. and Breen, A.C. (1999). The bournemouth questionnaire: A short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *J Manipulative Physiol Ther* 22 (8), 503-510.

27. Bolton, J.E. and Humphreys, B.K. (2002). The bournemouth questionnaire: A short-form comprehensive outcome measure. Ii. Psychometric properties in neck pain patients. *J Manipulative Physiol Ther* 25 (3), 141-148.
28. Bolton, P.S. (1998). The somatosensory system of the neck and its effects on the central nervous system. *J Manipulative Physiol Ther* 21 (8), 553-563.
29. Borghouts, J., Janssen, H., Koes, B., Muris, J., Metsemakers, J. and Bouter, L. (1999). The management of chronic neck pain in general practice. A retrospective study. *Scand J Prim Health Care* 17 (4), 215-220.
30. Bot, S.D., Van Der Waal, J.M., Terwee, C.B., Van Der Windt, D.A., Scholten, R.J., Bouter, L.M. and Dekker, J. (2005). Predictors of outcome in neck and shoulder symptoms: A cohort study in general practice. *Spine* 30 (16), E459-470.
31. Boyd-Clark, L.C., Briggs, C.A. and Galea, M.P. (2002). Muscle spindle distribution, morphology, and density in longus colli and multifidus muscles of the cervical spine. *Spine* 27 (7), 694-701.
32. Brandt, T. (1996). Cervical vertigo--reality or fiction? *Audiol Neurootol* 1 (4), 187-196.
33. Brandt, T. and Bronstein, A.M. (2001). Cervical vertigo. *J Neurol Neurosurg Psychiatry* 71 (1), 8-12.
34. Brault, J.R., Siegmund, G.P. and Wheeler, J.B. (2000). Cervical muscle response during whiplash: Evidence of a lengthening muscle contraction. *Clin Biomech (Bristol, Avon)* 15 (6), 426-435.
35. Brown, J.J. (1992). Cervical contribution to balance: Cervical vertigo. In: *The head-neck sensory motor system*. pp. 644-647. Berthoz, A., Vidal, P. and Graf, W. (eds). Oxford University Press, New York.
36. Bruton, A., Conway, J.H. and Holgate, S.T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy* 86 (2), 94-99.
37. Capuano-Pucci, D., Rheault, W., Aukai, J., Bracke, M., Day, R. and Pastrick, M. (1991). Intratester and intertester reliability of the cervical range of motion device. *Arch Phys Med Rehabil* 72 (5), 338-340.
38. Carroll, L., Hogg-Johnson, S., Cote, P., Van Der Velde, G., Holm, L.W., Carragee, E., Hurwitz, E.L., Peloso, P., Cassidy, J.D., Guzman, J., Nordin, M. and Haldeman, S. (2008). Course and prognostic factors for neck pain in workers: Results of the the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S93-S100.
39. Carroll, L., Holm, L.W., Hogg-Johnson, S., Cote, P., Cassidy, J.D., Haldeman, S., Nordin, M., Hurwitz, E.L., Carragee, E., Van Der Velde, G., Peloso, P. and Guzman, J. (2008). Course and prognostic factors for neck pain in whiplash-associated disorders (wad): Results of the the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S83-S92.
40. Carroll, L.J., Hogg-Johnson, S., Van Der Velde, G., Haldeman, S., Holm, L.W., Carragee, E., Hurwitz, E.L., Cote, P., Nordin, M., Peloso, P., Guzman, J. and Cassidy, J.D. (2008). Course and prognostic factors for neck pain in the general population: Results of the the bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S75-S82.
41. Ceyte, H., Cian, C., Nougier, V., Olivier, I. and Roux, A. (2006). Effects of neck muscles vibration on the perception of the head and trunk midline position. *Exp Brain Res* 170 (1), 136-140.
42. Chester, J.B., Jr. (1991). Whiplash, postural control, and the inner ear. *Spine* 16 (7), 716-720.
43. Chinn, S. (1990). The assessment of methods of measurement. *Stat Med* 9 (4), 351-362.
44. Chiu, T.T. and Leung, A.S. (2006). Neck pain in hong kong: A telephone survey on prevalence, consequences, and risk groups. *Spine* 31 (16), E540-544.
45. Christensen, H.W. and Nilsson, N. (1999). The ability to reproduce the neutral zero position of the head. *J Manipulative Physiol Ther* 22 (1), 26-28.
46. Cote, P., Cassidy, J.D., Carroll, L.J. and Kristman, V. (2004). The annual incidence and course of neck pain in the general population: A population-based cohort study. *Pain* 112 (3), 267-273.
47. De Hertogh, W., Vaes, P., Beckwee, D., Van Suijlekom, H., Duquet, W. and Van Roy, P. (2008). Lack of impairment of kinaesthetic sensibility in cervicogenic headache patients. *Cephalalgia* 28 (4), 323-328.
48. Demaille-Wlodyka, S., Chiquet, C., Lavaste, J.F., Skalli, W., Revel, M. and Poiraudreau, S. (2007). Cervical range of motion and cephalic kinesthesia: Ultrasonographic analysis by age and sex. *Spine* 32 (8), E254-261.
49. Devereaux, M.W. (2004). Neck pain. *Prim Care* 31 (1), 19-31.
50. Dutia, M.B. (1991). The muscles and joints of the neck: Their specialisation and role in head movement. *Prog Neurobiol* 37 (2), 165-178.
51. Dvir, Z. and Prushansky, T. (2000). Reproducibility and instrument validity of a new ultrasonography-based system for measuring cervical spine kinematics [in process citation]. *Clin Biomech (Bristol, Avon)* 15 (9), 658-664.

52. Dvir, Z., Prushansky, T. and Peretz, C. (2001a). Maximal versus feigned active cervical motion in healthy patients: The coefficient of variation as an indicator for sincerity of effort. *Spine* 26 (15), 1680-1688.
53. Dvir, Z., Prushansky, T. and Peretz, C. (2001b). Cervical circumduction in normal subjects: A study of maximal versus feigned effort. *J Spinal Disord* 14 (6), 472-478.
54. Ektor-Andersen, J., Isacsson, S.O., Lindgren, A. and Orbaek, P. (1999). The experience of pain from the shoulder-neck area related to the total body pain, self-experienced health and mental distress. The malmo shoulder-neck study group. *Pain* 82 (3), 289-295.
55. Eliasziw, M., Young, S.L., Woodbury, M.G. and Fryday-Field, K. (1994). Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Phys Ther* 74 (8), 777-788.
56. Ergen, E. and Ulkar, B. (2008). Proprioception and ankle injuries in soccer. *Clin Sports Med* 27 (1), 195-217, x.
57. Ernst, A., Basta, D., Seidl, R.O., Todt, I., Scherer, H. and Clarke, A. (2005). Management of posttraumatic vertigo. *Otolaryngol Head Neck Surg* 132 (4), 554-558.
58. Feipel, V., Salvia, P., Klein, H. and Rooze, M. (2005). Head repositioning accuracy in patients with whiplash-associated disorders. *Comput Methods Biomech Biomed Engin Supp* 1, 97-98.
59. Feipel, V., Salvia, P., Klein, H. and Rooze, M. (2006). Head repositioning accuracy in patients with whiplash-associated disorders. *Spine* 31 (2), E51-58.
60. Ferrari, R. and Russell, A.S. (2003). Regional musculoskeletal conditions: Neck pain. *Best Pract Res Clin Rheumatol* 17 (1), 57-70.
61. Field, A. (2005). *Discovering statistics using spss. p.^pp. Pages.* Sage Publications Ltd, London.
62. Fleiss, J.L. (1986). Reliability of measurement. In: *Design and analysis of clinical experiments.* Fleiss, J.L. (ed). John Wiley & Sons, New York.
63. Friedenber, Z.B. and Miller, W.T. (1963). Degenerative disc disease of the cervical spine: A comparative study of asymptomatic and symptomatic patients. *J Bone Joint Surg [Am]* 45, 1171-1178.
64. Gandevia, S.C., McCloskey, D.I. and Burke, D. (1992). Kinaesthetic signals and muscle contraction. *Trends Neurosci* 15 (2), 62-65.
65. Gimse, R., Tjell, C., Bjorgen, I.A. and Saunte, C. (1996). Disturbed eye movements after whiplash due to injuries to the posture control system. *J Clin Exp Neuropsychol* 18 (2), 178-186.
66. Goebel, J.A., Hanson, J.M. and Fishel, D.G. (1994). Age-related modulation of the vestibulo-ocular reflex using real and imaginary targets. *J Vestib Res* 4 (4), 269-275.
67. Gore, D.R., Sepic, S.B. and Gardner, G.M. (1986). Roentgenographic findings of the cervical spine in asymptomatic people. *Spine* 11 (6), 521-524.
68. Green, A.M., Shaikh, A.G. and Angelaki, D.E. (2005). Sensory vestibular contributions to constructing internal models of self-motion. *J Neural Eng* 2 (3), S164-179.
69. Guzman, J., Hurwitz, E.L., Carroll, L., Haldeman, S., Cote, P., Carragee, E., Peloso, P., Van Der Velde, G., Holm, L.W., Hogg-Johnson, S., Nordin, M. and Cassidy, J.D. (2008). A new conceptual model of neck pain - linking onset, course, and care: The bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S14-S23.
70. Haas, M. (1991a). The reliability of reliability. *J Manipulative Physiol Ther* 14 (3), 199-208.
71. Haas, M. (1991b). Statistical methodology for reliability studies. *J Manipulative Physiol Ther* 14 (2), 119-132.
72. Hagen, K., Einarsen, C., Zwart, J.A., Svebak, S. and Bovim, G. (2002). The co-occurrence of headache and musculoskeletal symptoms amongst 51 050 adults in norway. *Eur J Neurol* 9 (5), 527-533.
73. Haldeman, S., Carroll, L.J., Cassidy, J.D., Schubert, J. and Nygren, A. (2008). The bone and joint decade 2000-2010 task force on neck pain and its associated disorders: Executive summary. *Spine* 33 (4 Suppl), S5-S7.
74. Haldeman, S., Carroll, L.J. and Cassidy, J.D. (2008). The empowerment of people with neck pain: Introduction: The bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S8-S13.
75. Hallgren, R.C., Greenman, P.E. and Rechten, J.J. (1994). Atrophy of suboccipital muscles in patients with chronic pain: A pilot study. *J Am Osteopath Assoc* 94 (12), 1032-1038.
76. Hamann, K.F., Strauss, K., Kellner, M. and Weiss, U. (1992). Dependence of visual straight ahead on vestibular influences. XVIIth Barany Society Meeting. Castle Dobris, Czechoslovakia.
77. Hartling, L., Brison, R.J., Ardern, C. and Pickett, W. (2001). Prognostic value of the quebec classification of whiplash-associated disorders. *Spine* 26 (1), 36-41.
78. Harvey, V.L. and Dickenson, A.H. (2008). Mechanisms of pain in nonmalignant disease. *Curr Opin Support Palliat Care* 2 (2), 133-139.



79. Heikkila, H. and Astrom, P.G. (1996). Cervicocephalic kinesthetic sensibility in patients with whiplash injury. *Scand J Rehabil Med* 28 (3), 133-138.
80. Heikkila, H., Johansson, M. and Wenngren, B.I. (2000). Effects of acupuncture, cervical manipulation and nsaid therapy on dizziness and impaired head repositioning of suspected cervical origin: A pilot study. *Man Ther* 5 (3), 151-157.
81. Heikkila, H.V. and Wenngren, B.I. (1998). Cervicocephalic kinesthetic sensibility, active range of cervical motion, and oculomotor function in patients with whiplash injury. *Arch Phys Med Rehabil* 79 (9), 1089-1094.
82. Heller, C.A., Stanley, P., Lewis-Jones, B. and Heller, R.F. (1983). Value of x ray examinations of the cervical spine. *Br Med J (Clin Res Ed)* 287 (6401), 1276-1278.
83. Hill, J., Lewis, M., Papageorgiou, A.C., Dziedzic, K. and Croft, P. (2004). Predicting persistent neck pain: A 1-year follow-up of a population cohort. *Spine* 29 (15), 1648-1654.
84. Hogg-Johnson, S., Van Der Velde, G., Carroll, L., Holm, L.W., Cassidy, J.D., Guzman, J., Cote, P., Haldeman, S., Ammendolia, C., Carragee, E., Hurwitz, E.L., Nordin, M. and Peloso, P. (2008). The burden and determinants of neck pain in the general population: The bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S39-S51.
85. Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Med* 30 (1), 1-15.
86. Humphreys, B.K. and Irgens, P.M. (2002). The effect of a rehabilitation exercise program on head repositioning accuracy and reported levels of pain in chronic neck pain subjects. *Journal of Whiplash & Related Disorders* 1 (1), 99-112.
87. Hurley, M.V., Jones, D.W. and Newham, D.J. (1994). Arthrogenic quadriceps inhibition and rehabilitation of patients with extensive traumatic knee injuries [published erratum appears in *clin sci* 1994 jun;86(6):Xxii]. *Clin Sci (Colch)* 86 (3), 305-310.
88. Hurley, M.V., Scott, D.L., Rees, J. and Newham, D.J. (1997). Sensorimotor changes and functional performance in patients with knee osteoarthritis. *Ann Rheum Dis* 56 (11), 641-648.
89. Hurwitz, E.L., Coulter, I.D., Adams, A.H., Genovese, B.J. and Shekelle, P.G. (1998). Use of chiropractic services from 1985 through 1991 in the united states and canada. *Am J Public Health* 88 (5), 771-776.
90. Jaeschke, R., Guyatt, G.H. and Sackett, D.L. (1994). Users' guides to the medical literature. Iii. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The evidence-based medicine working group. *JAMA* 271 (9), 703-707.
91. Johansson, H., Djupsjobacka, M. and Sjolander, P. (1993). Influences on the gamma-muscle spindle system from muscle afferents stimulated by kcl and lactic acid. *Neurosci Res* 16 (1), 49-57.
92. Kandel, E.R., Schwartz, J.H. and Jessell, T.M. (2000). Principles of neural science. p.^pp. Pages. Fourth ed. McGraw-Hill
93. Karlsborg, M., Smed, A., Jespersen, H., Stephensen, S., Cortsen, M., Jennum, P., Herning, M., Korfitsen, E. and Werdelin, L. (1997). A prospective study of 39 patients with whiplash injury. *Acta Neurol Scand* 95 (2), 65-72.
94. Karnath, H.O., Sievering, D. and Fetter, M. (1994). The interactive contribution of neck muscle proprioception and vestibular stimulation to subjective "Straight ahead" Orientation in man. *Exp Brain Res* 101 (1), 140-146.
95. Kelders, W.P., Kleinrensink, G.J., Van Der Geest, J.N., Schipper, I.B., Feenstra, L., De Zeeuw, C.I. and Frens, M.A. (2005). The cervico-ocular reflex is increased in whiplash injury patients. *J Neurotrauma* 22 (1), 133-137.
96. Ketroser, D.B. (2000). Whiplash, chronic neck pain, and zygapophyseal joint disorders. A selective review. *Minn Med* 83 (2), 51-54.
97. Kibler, W.B. and Livingston, B. (2001). Closed-chain rehabilitation for upper and lower extremities. *J Am Acad Orthop Surg* 9 (6), 412-421.
98. Kim, K., Uchiyama, M., Liu, X., Shibui, K., Ohida, T., Ogihara, R. and Okawa, M. (2001). Somatic and psychological complaints and their correlates with insomnia in the japanese general population. *Psychosom Med* 63 (3), 441-446.
99. Kingma, H. (2006). Function tests of the otolith or statolith system. *Curr Opin Neurol* 19 (1), 21-25.
100. Kogler, A., Lindfors, J., Odkvist, L.M. and Ledin, T. (2000). Postural stability using different neck positions in normal subjects and patients with neck trauma. *Acta Otolaryngol* 120 (2), 151-155.
101. Koumantakis, G.A., Winstanley, J. and Oldham, J.A. (2002). Thoracolumbar proprioception in individuals with and without low back pain: Intratester reliability, clinical applicability, and validity. *J Orthop Sports Phys Ther* 32 (7), 327-335.
102. Kristjansson, E., Dall'alba, P. and Jull, G. (2001). Cervicocephalic kinaesthesia: Reliability of a new test approach. *Physiother Res Int* 6 (4), 224-235.

103. Kristjansson, E., Dall'alba, P. and Jull, G. (2003). A study of five cervicocephalic relocation tests in three different subject groups. *Clin Rehabil* 17 (7), 768-774.
104. Kristjansson, E., Hardardottir, L., Asmundardottir, M. and Gudmundsson, K. (2004). A new clinical test for cervicocephalic kinesthetic sensibility: "The fly". *Arch Phys Med Rehabil* 85 (3), 490-495.
105. Kulkarni, V., Chandy, M.J. and Babu, K.S. (2001). Quantitative study of muscle spindles in suboccipital muscles of human fetuses. *Neurol India* 49 (4), 355-359.
106. Lachin, J.M. (2004). The role of measurement reliability in clinical trials. *Clin Trials* 1 (6), 553-566.
107. Lackner, J.R. and Levine, M.S. (1979). Changes in apparent body orientation and sensory localization induced by vibration of postural muscles: Vibratory myesthetic illusions. *Aviat Space Environ Med* 50 (4), 346-354.
108. Lam, S.S., Jull, G. and Treleaven, J. (1999). Lumbar spine kinesthesia in patients with low back pain. *J Orthop Sports Phys Ther* 29 (5), 294-299.
109. Lang, T.A. and Secic, M. (1997). How to report statistics in medicine. p.^pp. Pages. American College of Physicians, Philadelphia.
110. Lang, T.A. and Secic, M. (2006). How to report statistics in medicine. p.^pp. Pages. Second ed. American College of Physicians, Philadelphia.
111. Lark, S.D. and Mccarthy, P.W. (2007). Cervical range of motion and proprioception in rugby players versus non-rugby players. *J Sports Sci* 25 (8), 887-894.
112. Lee, H., Nicholson, L.L., Adams, R.D. and Bae, S.S. (2005). Proprioception and rotation range sensitization associated with subclinical neck pain. *Spine* 30 (3), E60-67.
113. Lee, H.Y., Teng, C.C., Chai, H.M. and Wang, S.F. (2006). Test-retest reliability of cervicocephalic kinesthetic sensibility in three cardinal planes. *Man Ther* 11 (1), 61-68.
114. Lee, H.Y., Wang, J.D., Yao, G. and Wang, S.F. (2007). Association between cervicocephalic kinesthetic sensibility and frequency of subclinical neck pain. *Man Ther*.
115. Lee, J., Koh, D. and Ong, C.N. (1989). Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 19 (1), 61-70.
116. Lewit, K. (1994). The functional approach. *J Orthop Med* 16 (3), 73-74.
117. Lidgren, L. (2008). Preface: Neck pain and the decade of the bone and joint 2000-2010. *Spine* 33 (4 Suppl), S1-S2.
118. Liebenson, C.S. (2007). Rehabilitation of the spine - a practitioners manual. p.^pp. Pages. 2nd ed
119. Liu, J.X., Thornell, L.E. and Pedrosa-Domellof, F. (2003). Muscle spindles in the deep muscles of the human neck: A morphological and immunocytochemical study. *J Histochem Cytochem* 51 (2), 175-186.
120. Lonn, J., Crenshaw, A.G., Djupsjobacka, M. and Johansson, H. (2000). Reliability of position sense testing assessed with a fully automated system. *Clin Physiol* 20 (1), 30-37.
121. Lonn, J., Crenshaw, A.G., Djupsjobacka, M., Pedersen, J. and Johansson, H. (2000). Position sense testing: Influence of starting position and type of displacement. *Arch Phys Med Rehabil* 81 (5), 592-597.
122. Lord, S.M., Barnsley, L., Wallis, B.J. and Bogduk, N. (1996). Chronic cervical zygapophysial joint pain after whiplash. A placebo-controlled prevalence study. *Spine* 21 (15), 1737-1744; discussion 1744-1735.
123. Loudon, J.K., Ruhl, M. and Field, E. (1997). Ability to reproduce head position after whiplash injury. *Spine* 22 (8), 865-868.
124. Ludbrook, J. (1997). Comparing methods of measurements. *Clin Exp Pharmacol Physiol* 24 (2), 193-203.
125. Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clin Exp Pharmacol Physiol* 29 (7), 527-536.
126. Maffey-Ward, L., Jull, G. and Wellington, L. (1996). Toward a clinical test of lumbar spine kinesthesia. *J Orthop Sports Phys Ther* 24 (6), 354-358.
127. Maher, C. (1993). Pitfalls in reliability studies: Some suggestions for change. *Australian Physiotherapy* 39 (1), 5-7.
128. Mannion, A.F., Klein, G.N., Dvorak, J. and Lanz, C. (2000). Range of global motion of the cervical spine: Intraindividual reliability and the influence of measurement device. *Eur Spine J* 9 (5), 379-385.
129. Marchiori, D.M. and Henderson, C.N. (1996). A cross-sectional study correlating cervical radiographic degenerative findings to pain and disability. *Spine* 21 (23), 2747-2751.
130. Matre, D.A., Sinkjaer, T., Svensson, P. and Arendt-Nielsen, L. (1998). Experimental muscle pain increases the human stretch reflex. *Pain* 75 (2-3), 331-339.
131. McGraw, K.O. and Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1 (1), 30-46.
132. Mclain, R.F. (1994). Mechanoreceptor endings in human cervical facet joints. *Spine* 19 (5), 495-501.

133. McNair, P.J., Portero, P., Chiquet, C., Mawston, G. and Lavaste, F. (2006). Acute neck pain: Cervical spine range of motion and position sense prior to and after joint mobilization. *Man Ther.*
134. McPartland, J.M., Brodeur, R.R. and Hallgren, R.C. (1997). Chronic neck pain, standing balance, and suboccipital muscle atrophy--a pilot study. *J Manipulative Physiol Ther* 20 (1), 24-29.
135. McPartland, J.M. and Brodeur, R.R. (1999). Rectus capitis posterior minor: A small but important suboccipital muscle. *J Bodywork and Movement Therapies* 3 (1), 30-35.
136. Merfeld, D.M., Park, S., Gianna-Poulin, C., Black, F.O. and Wood, S. (2005). Vestibular perception and action employ qualitatively different mechanisms. I. Frequency response of vor and perceptual responses during translation and tilt. *J Neurophysiol* 94 (1), 186-198.
137. Mergner, T., Nardi, G.L., Becker, W. and Deecke, L. (1983). The role of canal-neck interaction for the perception of horizontal trunk and head rotation. *Exp Brain Res* 49 (2), 198-208.
138. Mergner, T., Siebold, C., Schweigart, G. and Becker, W. (1991). Human perception of horizontal trunk and head rotation in space during vestibular and neck stimulation. *Exp Brain Res* 85 (2), 389-404.
139. Mergner, T., Hlavacka, F. and Schweigart, G. (1993). Interaction of vestibular and proprioceptive inputs. *J Vestib Res* 3 (1), 41-57.
140. Mikkelsen, M., Salminen, J.J. and Kautiainen, H. (1997). Non-specific musculoskeletal pain in preadolescents. Prevalence and 1-year persistence. *Pain* 73 (1), 29-35.
141. Mikkelsen, M., Sourander, A., Salminen, J.J., Kautiainen, H. and Piha, J. (1999). Widespread pain and neck pain in schoolchildren. A prospective one-year follow-up study. *Acta Paediatr* 88 (10), 1119-1124.
142. Montfoort, I., Kelders, W.P., Van Der Geest, J.N., Schipper, I.B., Feenstra, L., De Zeeuw, C.I. and Frens, M.A. (2006). Interaction between ocular stabilization reflexes in patients with whiplash injury. *Invest Ophthalmol Vis Sci* 47 (7), 2881-2884.
143. Muller, R. and Buttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Stat Med* 13 (23-24), 2465-2476.
144. Murphy, D.R. (2000). Dysfunction in the cervical spine. In: *Conservative management of cervical spine syndromes*. pp. 71-103. Murphy, D.R. (ed). McGraw-Hill, New York.
145. Murphy, D.R. (2004). A clinical model for the diagnosis and management of patients with cervical spine syndromes *Australasian Journal of Chiropractic and Osteopathy* 12 (2), 57-71.
146. Nitz, A.J. and Peck, D. (1986). Comparison of muscle spindle concentrations in large and small human epaxial muscles acting in parallel combinations. *Am Surg* 52 (5), 273-277.
147. O'sullivan, P. (2005). Diagnosis and classification of chronic low back pain disorders: Maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 10 (4), 242-255.
148. Ottenbacher, K.J. (1995). An examination of reliability in developmental research. *J Dev Behav Pediatr* 16 (3), 177-182.
149. Palmer, K.T., Syddall, H., Cooper, C. and Coggon, D. (2003). Smoking and musculoskeletal disorders: Findings from a british national survey. *Ann Rheum Dis* 62 (1), 33-36.
150. Palmgren, P.J., Sandstrom, P.J., Lundqvist, F.J. and Heikkila, H. (2006). Improvement after chiropractic care in cervicocephalic kinesthetic sensibility and subjective pain intensity in patients with nontraumatic chronic neck pain. *J Manipulative Physiol Ther* 29 (2), 100-106.
151. Panjabi, M.M. (1992a). The stabilizing system of the spine. Part ii. Neutral zone and instability hypothesis. *J Spinal Disord* 5 (4), 390-396; discussion 397.
152. Panjabi, M.M. (1992b). The stabilizing system of the spine. Part i. Function, dysfunction, adaptation, and enhancement. *J Spinal Disord* 5 (4), 383-389; discussion 397.
153. Panjabi, M.M. (2006). A hypothesis of chronic back pain: Ligament subfailure injuries lead to muscle control dysfunction. *Eur Spine J* 15 (5), 668-676.
154. Peck, D., Buxton, D.F. and Nitz, A. (1984). A comparison of spindle concentrations in large and small muscles acting in parallel combinations. *J Morphol* 180 (3), 243-252.
155. Pedersen, P., Noddeskou, H. and Wejse, B. (1992). A pilot survey of diagnoses, radiographic and laboratory procedures encountered in european chiropractic practices. *European Journal of Chiropractic* 40, 71-82.
156. Pernold, G., Mortimer, M., Wiktorin, C., Tornqvist, E.W. and Vingard, E. (2005). Neck/shoulder disorders in a general population. Natural course and influence of physical exercise: A 5-year follow-up. *Spine* 30 (13), E363-368.
157. Peterson, C., Bolton, J., Wood, A.R. and Humphreys, B.K. (2003). A cross-sectional study correlating degeneration of the cervical spine with disability and pain in united kingdom patients. *Spine* 28 (2), 129-133.
158. Pettersson, K., Hildingsson, C., Toolanen, G., Fagerlund, M. and Bjornebrink, J. (1994). Mri and neurology in acute whiplash trauma. No correlation in prospective examination of 39 cases. *Acta Orthop Scand* 65 (5), 525-528.

159. Proske, U., Schaible, H.G. and Schmidt, R.F. (1988). Joint receptors and kinaesthesia. *Exp Brain Res* 72 (2), 219-224.
160. Quan, H. and Shih, W.J. (1996). Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics* 52 (4), 1195-1203.
161. Rankin, G. and Stokes, M. (1998). Reliability of assessment tools in rehabilitation: An illustration of appropriate statistical analyses. *Clin Rehabil* 12 (3), 187-199.
162. Revel, M., Andre-Deshays, C. and Minguet, M. (1991). Cervicocephalic kinesthetic sensibility in patients with cervical pain. *Arch Phys Med Rehabil* 72 (5), 288-291.
163. Revel, M., Minguet, M., Gergoy, P., Vaillant, J. and Manuel, J.L. (1994). Changes in cervicocephalic kinesthesia after a proprioceptive rehabilitation program in patients with neck pain: A randomized controlled study. *Arch Phys Med Rehabil* 75 (8), 895-899.
164. Richmond, F. (1988). The sensorium: Receptors of neck muscles and joints. In: *Control of head movement*. pp. 49 - 62. Peterson, B.W. and Richmond, F.J. (eds). Oxford University Press, New York.
165. Rogers, R.G. (1997). The effects of spinal manipulation on cervical kinesthesia in patients with chronic neck pain: A pilot study. *J Manipulative Physiol Ther* 20 (2), 80-85.
166. Ronnen, H.R., De Korte, P.J., Brink, P.R., Van Der Bijl, H.J., Tonino, A.J. and Franke, C.L. (1996). Acute whiplash injury: Is there a role for mr imaging?--a prospective study of 100 patients. *Radiology* 201 (1), 93-96.
167. Rydevik, B. (2008). Editorial preface: The bone and joint decade 2000-2010 task force on neck pain and its associated disorders. *Spine* 33 (4 Suppl), S3.
168. Schaible, H.G. and Grubb, B.D. (1993). Afferent and spinal mechanisms of joint pain. *Pain* 55 (1), 5-54.
169. Schappert, S.M. (1996). National ambulatory medical care survey: 1994 summary. *Adv Data* (273), 1-18.
170. Schuck, P. (2004). Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Qual Life Res* 13 (3), 571-586.
171. Seaman, D.R. and Winterstein, J.F. (1998). Dysafferentation: A novel term to describe the neuropathophysiological effects of joint complex dysfunction. A look at likely mechanisms of symptom generation. *J Manipulative Physiol Ther* 21 (4), 267-280.
172. Shekelle, P.G. and Brook, R.H. (1991). A community-based study of the use of chiropractic services. *Am J Public Health* 81 (4), 439-442.
173. Sherrington, C.S. (1906). On the proprioceptive system, especially in its reflex aspects. *Brain* 29, 467-482.
174. Shrout, P.E. and Fleiss, J.L. (1979). Intraclass correlations; uses in assessing rater reliability. *Psychol Bull* 86, 420-428.
175. Sjolander, P., Michaelson, P., Jaric, S. and Djupsjobacka, M. (2008). Sensorimotor disturbances in chronic neck pain--range of motion, peak velocity, smoothness of movement, and repositioning acuity. *Man Ther* 13 (2), 122-131.
176. Soderlund, A., Olerud, C. and Lindberg, P. (2000). Acute whiplash-associated disorders (wad): The effects of early mobilization and prognostic factors in long-term symptomatology. *Clin Rehabil* 14 (5), 457-467.
177. Spitzer, W.O., Skovron, M.L., Salmi, L.R., Cassidy, J.D., Duranceau, J., Suissa, S. and Zeiss, E. (1995). Scientific monograph of the quebec task force on whiplash-associated disorders: Redefining "Whiplash" And its management [see comments] [published erratum appears in spine 1995 nov 1;20(21):2372]. *Spine* 20 (8 Suppl), 1S-73S.
178. Sterling, M., Jull, G., Vicenzino, B., Kenardy, J. and Darnell, R. (2003). Development of motor system dysfunction following whiplash injury. *Pain* 103 (1-2), 65-73.
179. Sterling, M. (2004). A proposed new classification system for whiplash associated disorders--implications for assessment and management. *Man Ther* 9 (2), 60-70.
180. Strimpakos, N., Sakellari, V., Giftofos, G., Kapreli, E. and Oldham, J. (2006). Cervical joint position sense: An intra- and inter-examiner reliability study. *Gait and Posture* 23 (1), 22-31.
181. Strupp, M., Arbusow, V., Borges Pereira, C., Dieterich, M. and Brandt, T. (1999). Subjective straight-ahead during neck muscle vibration: Effects of ageing. *Neuroreport* 10 (15), 3191-3194.
182. Stussman, B.J. (1996). National hospital ambulatory medical care survey: 1993 emergency department summary. *Adv Data* (271), 1-15.
183. Swait, G., Rushton, A.B., Miall, R.C. and Newell, D. (2007). Evaluation of cervical proprioceptive function: Optimizing protocols and comparison between tests in normal subjects. *Spine* 32 (24), E692-701.
184. Swinkels, A. and Dolan, P. (2000). Spinal position sense is independent of the magnitude of movement. *Spine* 25 (1), 98-104; discussion 105.

185. Taylor, J.L. and McCloskey, D.I. (1988). Proprioception in the neck. *Exp Brain Res* 70 (2), 351-360.
186. Teng, C.C., Chai, H., Lai, D.M. and Wang, S.F. (2006). Cervicocephalic kinesthetic sensibility in young and middle-aged adults with or without a history of mild neck pain. *Man Ther*.
187. Teng, C.C., Chai, H., Lai, D.M. and Wang, S.F. (2007). Cervicocephalic kinesthetic sensibility in young and middle-aged adults with or without a history of mild neck pain. *Man Ther* 12 (1), 22-28.
188. Thunberg, J., Hellstrom, F., Sjolander, P., Bergenheim, M., Wenngren, B. and Johansson, H. (2001). Influences on the fusimotor-muscle spindle system from chemosensitive nerve endings in cervical facet joints in the cat: Possible implications for whiplash induced disorders. *Pain* 91 (1-2), 15-22.
189. Thunberg, J., Ljubisavljevic, M., Djupsjobacka, M. and Johansson, H. (2002). Effects on the fusimotor-muscle spindle system induced by intramuscular injections of hypertonic saline. *Exp Brain Res* 142 (3), 319-326.
190. Treleaven, J., Jull, G. and Sterling, M. (2003). Dizziness and unsteadiness following whiplash injury: Characteristic features and relationship with cervical joint position error. *J Rehabil Med* 35 (1), 36-43.
191. Treleaven, J., Jull, G. and Lowchoy, N. (2006). The relationship of cervical joint position error to balance and eye movement disturbances in persistent whiplash. *Man Ther* 11 (2), 99-106.
192. Treleaven, J., Lowchoy, N., Darnell, R., Panizza, B., Brown-Rothwell, D. and Jull, G. (2008). Comparison of sensorimotor disturbance between subjects with persistent whiplash-associated disorder and subjects with vestibular pathology associated with acoustic neuroma. *Arch Phys Med Rehabil* 89 (3), 522-530.
193. Uhlig, Y., Weber, B.R., Grob, D. and Muntener, M. (1995). Fiber composition and fiber transformations in neck muscles of patients with dysfunction of the cervical spine. *J Orthop Res* 13 (2), 240-249.
194. Vadivelu, N. and Sinatra, R. (2005). Recent advances in elucidating pain mechanisms. *Curr Opin Anaesthesiol* 18 (5), 540-547.
195. Wall, P.D. (1989). The dorsal horn. In: *Textbook of pain*. pp. 102-111. Wall, P.D. and Melzack, R. (eds). Churchill Livingstone, Edinburgh.
196. Weber, B.R., Uhlig, Y., Grob, D., Dvorak, J. and Muntener, M. (1993). Duration of pain and muscular adaptations in patients with dysfunction of the cervical spine. *J Orthop Res* 11 (6), 805-810.
197. Weir, J.P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the sem. *J Strength Cond Res* 19 (1), 231-240.
198. Westerling, D. and Jonsson, B.G. (1980). Pain from the neck-shoulder region and sick leave. *Scand J Soc Med* 8 (3), 131-136.
199. Westgard, J.O. and Hunt, M.R. (1973). Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 19 (1), 49-57.
200. White, S.A. and Van Den Broek, N.R. (2004). Methods for assessing reliability and validity for a measurement tool: A case study and critique using the who haemoglobin colour scale. *Stat Med* 23 (10), 1603-1619.
201. Wilson, V.J. (1992). Physiologic properties and central actions of neck muscle spindles. In: *The head-neck sensory motor system*. pp. 175-178. Berthoz, A., Vidal, P. and Graf, W. (eds). Oxford University Press, New York.
202. Wrisley, D.M., Sparto, P.J., Whitney, S.L. and Furman, J.M. (2000). Cervicogenic dizziness: A review of diagnosis and treatment. *J Orthop Sports Phys Ther* 30 (12), 755-766.
203. Wuyts, F.L., Furman, J., Vanspauwen, R. and Van De Heyning, P. (2007). Vestibular function testing. *Curr Opin Neurol* 20 (1), 19-24.
204. Youdas, J.W., Carey, J.R. and Garrett, T.R. (1991). Reliability of measurements of cervical spine range of motion-- comparison of three methods. *Phys Ther* 71 (2), 98-104; discussion 105-106.

# Appendix I

**X BQ (Px) (PRE-TREATMENT-SCREENING VISIT-modified predictor Q) Patient reference number (for clinic use only):**

**X**

**This questionnaire is about your pain complaint. The information will be treated in complete confidence. For EACH question, TICK ONE box only unless instructed otherwise.**

Q1 PATIENT: START HERE: YOUR SURNAME:

Q2 First name:

Q3 Address (include postcode):


Q4 TODAY'S date:

Q5 Age (years):

**X**

Q6 Are you?

Male..... Female.....

Q7 What place(s) do you feel most pain?  
(more than one box allowed)

Low Back  Neck..... Shoulder-  
/arm.....  
Leg..... Head..... Other.....

Q8 If your pain is in your back or neck, does it go down into your leg(s) or your arm(s)?

Yes..... No.....

Q9 Would you describe your pain as generally 'ALL OVER' your body?

Yes..... No.....

Q10 Is your painful complaint the result of a specific injury/trauma?

Yes..... No..... Don't know.....

Q11 Have you had this same or a similar complaint in the past?

Yes..... No.....

Q12 How long has this PRESENT episode of your painful complaint lasted?

Less than 7 weeks. 7 weeks or longer..

Q13 How often are you taking medication (painkillers and other drugs) for your complaint?

A lot of the time  Occasionally/never.....

Q14 How do you expect your condition to change in the next few weeks?

Recover/improve..... Stay about the same.....  
Get worse.....

Q15 What is your current work status (tick ONE box only)?

Employed..... Retired.....  
Seeking work..... Working in the home.....

Q16 Are you overall satisfied with your current work status?

Yes..... No.....

Q17 In your estimation, do you expect to be working NORMALLY in 6 months time?

Yes/probably..... No/probably not....

Q18 Have you ever smoked?

Yes..... No.....

Q19 Do you drink alcohol?

Weekly..... Never/hardly ever.

Q20 Compared with people of a similar age and in a similar position, how would you rate your level of physical activity?

More/about the same..... Less.....

Q21 Apart from your complaint, how would you rate your GENERAL health and WELL-BEING?

Excellent/Good..... Fair/Poor.....

**X**

**CONTINUED OVERLEAF**

**X**

**X Put a CROSS in ONE box for EACH of the following statements that best describes your painful complaint and how it is affecting you NOW. Please read each question carefully before answering. X**

**Q22** Over the past few days, on average, how would you rate your pain on a scale where '0' is 'no pain' and '10' is 'worst pain possible'?

0      1      2      3      4      5      6      7      8      9      10

No pain

**Q23** Over the past few days, on average, how has your complaint interfered with your daily activities (housework, washing, dressing, lifting, walking, reading, driving, climbing stairs, getting in/out of bed/chair, sleeping) on a scale where '0' is 'no interference' and '10' is 'completely unable to carry on with normal daily activities'?

0      1      2      3      4      5      6      7      8      9      10

No interference

**Q24** Over the past few days, on average, how much has your painful complaint interfered with your normal social routine including recreational, social and family activities, on a scale where '0' is 'no interference' and '10' is 'completely unable to participate in any social and recreational activity'? X

0      1      2      3      4      5      6      7      8      9      10

No interference

**Q25** Over the past few days, on average, how anxious (uptight, tense, irritable, difficulty in relaxing/concentrating) have you been feeling, on a scale where '0' is 'not at all anxious' and '10' is 'extremely anxious'?

0      1      2      3      4      5      6      7      8      9      10

Not at all anxious

**Q26** Over the past few days, how depressed (down-in-the-dumps, sad, in low spirits, pessimistic, lethargic) have you been feeling, on a scale where '0' is 'not at all depressed' and '10' is 'extremely depressed'?

0      1      2      3      4      5      6      7      8      9      10

Not at all depressed

**Q27** Over the past few days, how do you think your work (both inside the home and/or employed work) have affected your painful complaint, on a scale where '0' is 'make it no worse' and '10' is 'make it very much worse'?

0      1      2      3      4      5      6      7      8      9      10

Make it no worse

**Q28** Over the past few days, on average, how much have you been able to control (help/reduce) and cope with your pain on your own, on a scale where '0' is 'I can control it completely' and '10' is 'I have no control whatsoever'?

0      1      2      3      4      5      6      7      8      9      10

I have complete control over my pain

**X THANK YOU VERY MUCH FOR YOUR TIME IN COMPLETING THIS QUESTIONNAIRE X**

Modified Nov 2000 (Predictor Q) Copyright AECC. All rights reserved. Contact Dr JE Bolton.

## Appendix II

Table 67. Raw artificial data for agreement statistics examples

X	X <sub>1</sub>	Y	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>
1.9	2.6	2.3	3.3	3.5	4.5	1.8	1
3.0	2.4	2.9	3.9	4.4	5.4	2.9	1.2
4.9	3.5	4.6	5.6	6.9	7.9	4.5	1.1
3.0	3.3	3.1	4.1	4.7	5.7	2.6	1.2
2.7	3.1	2.9	3.9	4.4	5.4	3.0	1.3
2.5	3.0	2.3	3.3	3.5	4.5	2.5	1.6
1.4	1.6	1.2	2.2	1.8	2.8	1.3	1.4
3.4	3.2	3.7	4.7	5.6	6.6	4.1	1.9
1.5	1.5	1.5	2.5	2.3	3.3	1.5	1.9
2.7	2.4	2.7	3.7	4.1	5.1	2.7	1.6
1.7	1.4	2.0	3.0	3.0	4.0	2.4	2.2
2.0	2.1	2.0	3.0	3.0	4.0	1.9	2.2
1.7	0.8	1.8	2.8	2.7	3.7	1.2	2
4.5	4.2	4.7	5.7	7.1	8.1	4.5	2.3
2.2	2.1	2.3	3.3	3.5	4.5	2.4	1.6
1.2	0.2	1.1	2.1	1.7	2.7	0.9	2
3.3	3.2	3.3	4.3	5.0	6.0	3.0	2.6
3.8	3.8	3.5	4.5	5.3	6.3	3.1	2.7
2.9	2.4	2.6	3.6	3.9	4.9	2.8	2.2
1.1	1.0	0.9	1.9	1.4	2.4	1.6	1.6
2.7	3.3	2.8	3.8	4.2	5.2	2.7	2
2.6	2.4	2.9	3.9	4.4	5.4	2.8	3.2
3.5	3.6	3.6	4.6	5.4	6.4	3.0	2.7
3.2	2.7	3.2	4.2	4.8	5.8	3.4	3.6
1.9	1.5	1.9	2.9	2.9	3.9	2.2	3.8
4.1	4.3	4.4	5.4	6.6	7.6	3.6	3.5
3.0	3.0	3.0	4.0	4.5	5.5	3.8	3.5
2.0	1.8	1.8	2.8	2.7	3.7	2.3	2.2
1.7	1.9	1.8	2.8	2.7	3.7	2.2	4
1.6	1.1	1.3	2.3	2.0	3.0	1.7	2.9
2.5	2.9	2.3	3.3	3.5	4.5	2.9	2.5
2.8	2.4	2.8	3.8	4.2	5.2	3.1	1.8
2.4	2.6	2.7	3.7	4.1	5.1	2.5	3.5
1.0	2.2	0.9	1.9	1.4	2.4	1.6	1.9
2.0	1.2	1.7	2.7	2.6	3.6	1.8	3.3
2.3	3.6	2.3	3.3	3.5	4.5	2.7	3.6
2.4	2.9	2.4	3.4	3.6	4.6	2.2	3
3.7	3.2	3.4	4.4	5.1	6.1	3.0	2.9
2.4	2.6	2.7	3.7	4.1	5.1	3.1	3.8
2.8	2.7	2.6	3.6	3.9	4.9	3.1	4.6
3.4	3.5	3.3	4.3	5.0	6.0	3.8	2.7
3.3	3.9	3.2	4.2	4.8	5.8	3.6	3.8
3.1	3.4	3.1	4.1	4.7	5.7	3.2	3.1
3.1	3.8	3.3	4.3	5.0	6.0	3.6	2
3.5	3.3	3.7	4.7	5.6	6.6	3.5	3
2.5	2.7	2.4	3.4	3.6	4.6	2.3	3
2.7	2.7	2.6	3.6	3.9	4.9	2.7	4
2.2	1.5	2.1	3.1	3.2	4.2	0.9	3.9
1.2	1.6	1.2	2.2	1.8	2.8	1.1	2.7
2.8	2.8	2.7	3.7	4.1	5.1	2.7	4.3



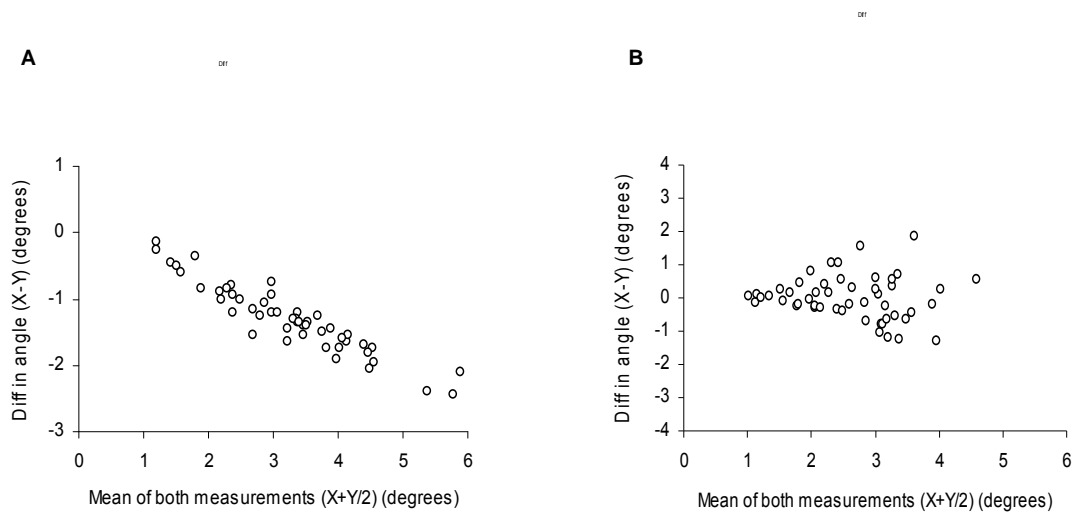
## Appendix III

### Comparison of methods using the Bland and Altman approach

For each subject, the difference between HRA-SSA values obtained using the two different measurement methods (5 reps – 10 reps mean for each subject) and the mean of these paired values ( $[(5 \text{ reps} - 10 \text{ reps})/2]$ ) were calculated. The mean difference ( $d$ ) and the standard deviation of the differences ( $s_d$ ) were then determined. Before calculating the limits of agreement, two assumptions were checked graphically and statistically:

- 1) The differences (5 reps – 10 reps) should come from an approximately Normal distribution. This was checked using a combination of statistical testing (Kolmogorov-Smirnov or K-S test) and observation of the data using frequency histograms, Q-Q (normal) plots and box plot summaries of the measures of central tendency (mean and median) and distribution (interquartile range).
- 2) The mean ( $d$ ) and standard deviation of the differences ( $s_d$ ) should be constant throughout the range of measurement. Put another way, there should be no relationship between the mean values and the differences. This was first checked visually using a scatter diagram (distribution plot) of the difference (5 reps – 10 reps) between the two measurement methods for each subject against their mean ( $[(5 \text{ reps} - 10 \text{ reps})/2]$ ). *Proportional bias* and *heteroscedasticity (proportional error)* are two common relationships between the differences and the magnitude that can occur in method agreement studies (Figure 61).

The possible presence of *proportional bias* was checked statistically using Pearson's product-moment correlation (assuming a normal distribution for both data means and differences). A *proportional bias* would be indicated by a statistically significant association between the difference and mean values ( $r$  significantly different from zero;  $P < 0.05$ ). An alternative approach is to perform an ordinary least squares (OLS) regression of the differences on the means. A *proportional bias* would be indicated by a slope of the line of best fit that is significantly different from zero (Ludbrook, 1997; Ludbrook, 2002).



**Figure 61.** Bland and Altman distribution plots with data exhibiting a relationship between the mean and difference: A – proportional bias; B – heteroscedasticity

The presence of *heteroscedasticity* can be further evaluated by calculating the absolute difference between the mean difference for the data set ( $\bar{d}$ ) and the observed differences between the methods for each subject. In essence this is the equivalent of the absolute residuals of a horizontal regression line through the original distribution plot. The absolute differences for each subject were then plotted against the mean of the values obtained from each method ( $[5 \text{ reps} - 10 \text{ reps}]/2$ ). As with the original distribution plots, an increase in the magnitude of the differences as the mean scores increase suggests *heteroscedasticity*. The presence of a *heteroscedasticity* was checked statistically by using Pearson's product-moment correlation.

*Heteroscedasticity* would be indicated by a statistically significant association between the absolute differences (ABS [observed differences-mean difference]) and mean values ( $r$  significantly different from zero;  $P < 0.05$ ). Lines representing the 95% limits of agreement and the mean difference were added to the scatter plot (see following for calculation of LoA). If the differences are normally distributed then 95% of the differences would be expected to lie between the limits of agreement. In addition, if no relationship exists within the data, the 5% of data points that sit outside the LoA lines should be approximately symmetrical in distribution above the upper LoA line, below the lower LoA line and throughout the range of measurements.

If the assumptions for the Bland and Altman analysis are not met, then the data should be log transformed before analysis or the limits of agreement calculated using a regression approach (Bland and Altman, 1999). For this study, the regression approach was utilised. This is described in a following section. If the assumptions are met, then the mean difference ( $d$ ) can be used directly as a measure of *fixed bias* between the methods of measurement. The 95% limits of agreement were then calculated from the mean difference  $\pm 1.96$  standard deviations of the differences (i.e.  $d \pm 1.96s_d$ ). The mean difference ( $d$ ) and the 95% LoA are only estimates for the population under investigation. The confidence intervals (CI) for  $d$  and the LoA allow some extrapolation of the results to other similar populations. These were calculated from the standard error (SE) of the respective measurements as follows:

Difference ( $d$ )

$$SE(d) = s_d/\sqrt{n}.$$

$$95\% \text{ CI for } d = d \pm tSE(d).$$

(value for  $t$  was determined by finding the appropriate point of the  $t$  distribution with  $n-1$  degrees of freedom)

Limits of Agreement (LoA)

$$SE(\text{LoA}) = 1.71SE(d)$$

$$95\% \text{ CI for LoA} = \text{upper LoA} \pm 1.96SE(\text{LoA})$$

$$\text{lower LoA} \pm 1.96SE(\text{LoA})$$

### ***Method agreement with repeated measurements – equal numbers of replicates***

When using repeated measurements instead of single measurements for two different methods of measurement (Figure 28B), the estimate of *fixed bias* (i.e. the mean difference,  $d$ ) will be unaffected but the estimate of the standard deviation of the differences ( $s_d$ ) will be too small because some of the effect of repeated measurement error has been removed. To correct for this, the adjusted (or corrected) standard deviation of the differences ( $s_c$ ) was calculated as follows:

$$s_c = \sqrt{s_d^2 + (s_{M1w}^2/2) + (s_{M2w}^2/2)} \quad 3.4$$

where:

$s_d^2$  (the square of the standard deviation of the differences) represents the observed variance of the differences between the test and retest mean values

$s_{M1w}^2$  &  $s_{M2w}^2$  (the mean of the squares of the standard deviations of the test and retest measurements for each measurement method; M1 and M2) represents the within-subject variance of repeated measurements for each measurement method.

In this case M1 = 5 reps; M2 = 10 reps.

The 95% limits of agreement were calculated from the mean difference  $\pm 1.96$  adjusted standard deviations (i.e.  $d \pm 1.96s_c$ ). An approximate standard error for these limits of agreement was calculated as follows.

$$SE(LoA) = \sqrt{s_c^2/n + 1.96^2/2s_c^2 (s_d^4/n-1 + s_{M1w}^4/4n + s_{M2w}^4/4n)} \quad 3.5$$

The 95% CI for LoA are calculated as detailed earlier using this adjusted standard error measurement.

***Method agreement with a relationship between difference and mean (magnitude)***

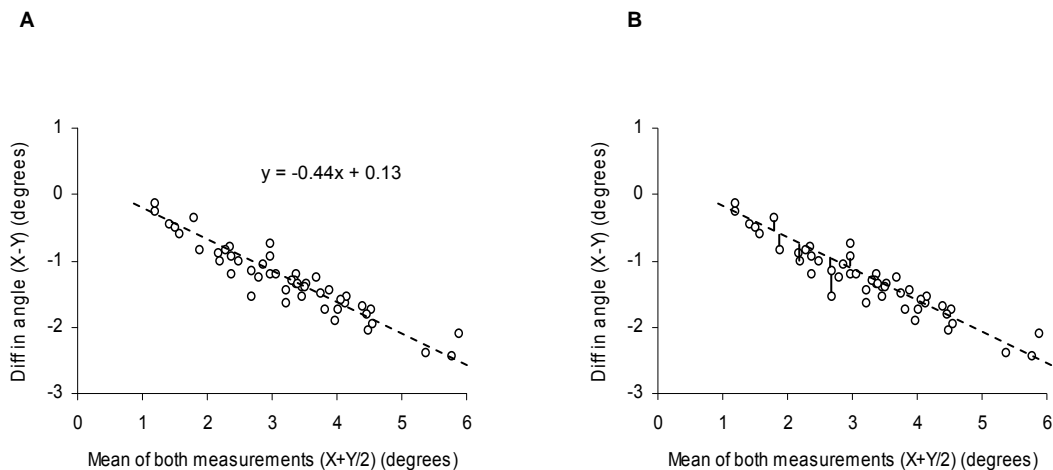
The previously described analysis assumes that the mean ( $d$ ) and standard deviation of the differences ( $s_d$ ) are the same throughout the range of measurement. Where relationships such as *proportional bias* or *heteroscedasticity* exist for the data (Figure 61), the LoA will be misleadingly wide when a *proportional bias* is present and will be too wide for the measurements in the low range and too narrow for the higher values for *heteroscedastic* data. Although these limits of agreement could be viewed as conservative and should not lead to the acceptance of poor methods of agreement they are nevertheless a crude estimate of the true agreement and somewhat misleading. In addition, as the estimate of *fixed bias* is

not independent of proportional bias using the standard Bland and Altman approach, inaccurate estimates of the fixed bias may occur in the presence of a *proportional bias* without a modified approach (Ludbrook, 1997; Bland and Altman, 1999; Ludbrook, 2002). For dealing with any relationship between  $d$ ,  $s_d$  and the magnitude of measurements, Bland and Altman describe a modification of their original approach using simple linear regression (Bland and Altman, 1999).

The analysis starts by simple regression (OLS) of difference of the methods (D) on the average (A) of the two methods giving the following equation:

$$\hat{D} = b_1A + b_0 \quad 3.6$$

The slope of the line of best-fit ( $b_1$ ) provides an index of the *proportional bias* between the methods and allows an estimate of the difference between methods for any value of the measurement. The intercept ( $b_0$ ) provides an estimate of the fixed bias between the methods (Figure 62A). If the slope is not significantly different from zero, such as with *heteroscedasticity* without *proportional bias*, then  $D = d$ , the mean difference.



**Figure 62.** A – Regression of difference of methods on the mean of two methods. B – Scatterplot to graphically define residuals; this is reflected by the vertical lines from the line of best fit to the observed measurements (NB; for clarity, not all lines are represented)

The relationship between the mean difference ( $d$ ) and the standard deviation of the differences ( $s_d$ ) can then be modelled using equation 3.1. The residuals about this line are obtained by taking the differences between the observed difference and the difference predicted by the regression for each of the mean measurements across the range (Figure 62B). The absolute residuals ( $R$ ) are then obtained (residuals without sign) and these values are then regressed on the average of the two methods ( $A$ ). This gives the following equation:

$$\hat{R} = c_1A + c_0 \quad 3.7$$

Assuming these residuals are normally distributed, multiplying these coefficients by  $\sqrt{\pi}/2$  gives an equation to predict the standard deviation of these differences ( $s_R$ ) for any value of the measurements. This is shown as follows:

$$s_R = \sqrt{\pi}/2 (c_1A + c_0)$$

or as

$$s_R = 1.25(c_1A + c_0) \quad 3.8$$

The modified limits of agreement are obtained by combining regression equations 3.6 and 3.8 as follows:

$$\text{LoA} = d \pm 1.96s_d$$

$$\text{Mod LoA} = \text{Predicted mean diffs} \pm 1.96 \text{ standard dev of predicted diffs}$$

$$= \text{Equ 3.6} \pm 1.96 \times \text{Equ 3.8}$$

$$= \hat{D} \pm 1.96(1.25\hat{R})$$

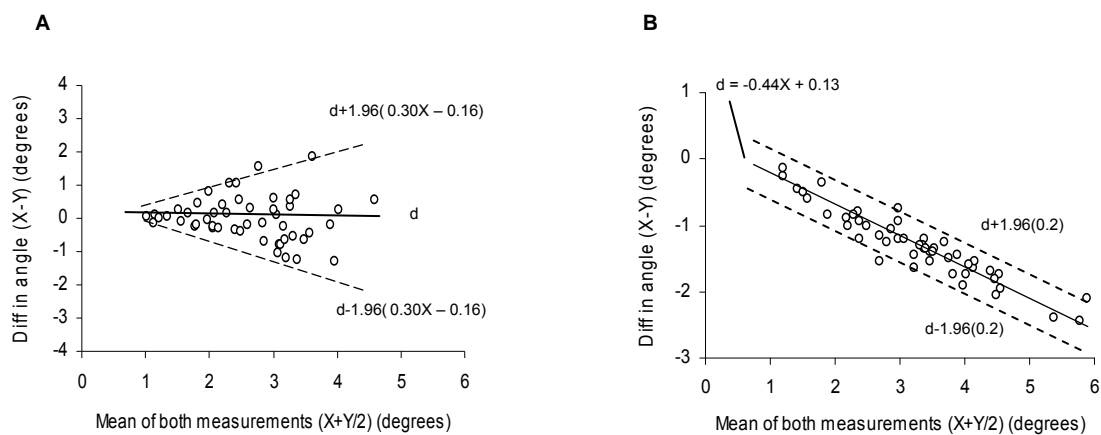
$$= b_1A + b_0 \pm 1.96(1.25(c_1A + c_0))$$

As stated earlier, if there is no *proportional bias*, only *heteroscedasticity*, then  $D = d$  (Figure 63A). Hence the equation for the modified LoA becomes:

$$= d \pm 1.96(1.25(c_1A + c_0))$$

If only *proportional bias* is exhibited, the variance of the absolute residuals will remain the same throughout the range of measurements and hence no significant predictive relationship will be seen when the residuals ( $R$ ) are regressed on the average of the two methods ( $A$ ); the slope of the line of best fit will not be significantly different from zero. In this instance, the standard deviation of these differences ( $s_R$ ) is simply the residual standard deviation from the regression line fitted through the original distribution plot. Therefore the equation for the modified LoA is:

$$\begin{aligned} \text{Mod LoA} &= \text{Predicted mean diffs} \pm 1.96 \text{ standard dev of predicted diffs} \\ &= \text{Equ 3.3} \pm 1.96 \times \text{standard deviation of the residuals} \\ &= b_1A + b_0 \pm 1.96(\text{standard deviation of the residuals}) \end{aligned}$$



**Figure 63.** A – Modified limits of agreement with equations for heteroscedastic data with no proportional bias ( $D = d$ ). B – Modified limits of agreement with equations for data exhibiting proportional bias but no heteroscedasticity.

As the standard deviation of the residuals is a constant value, the LoA lines will run parallel to the predicted differences line of best fit (Equation 3.3). This is shown in Figure 63B.

## Appendix IV

### **Study Title: Inter & Intra Examiner Reliability and Concurrent validity of the Zebris CMS 70P and Laser Pointer Methods for Measuring Cervicocephalic Kinaesthetic Sensibility**

Study Centre: Dept of Academic Affairs, Anglo-European College of Chiropractic (AECC),  
13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF. Tel: (01202) 436200

#### 1. Introduction

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and ask the investigator if there is anything that is not clear or if you would like more information.

Consumers for Ethics in Research (CERES) publish a leaflet entitled "Medical Research and You". This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained from CERES PO BOX 1365, London N10 0BW

#### 2. Purpose of the Study

The study will involve collecting information on the control of voluntary head movements. It is hoped that this information will shed more light on the usefulness of these tests described in section 5.

#### 3. Why have I been chosen?

You have been chosen because you do not suffer from neck pain and are otherwise healthy.

#### 4. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part you are still free to withdraw at anytime and without giving a reason.

**PLEASE TURN OVER**



5. What will happen to me if I take part?

You will be asked to attend the AECC outpatient clinic for a series of assessment procedures. These are as follows:

- Complete one short questionnaire regarding the eligibility criteria.
- Perform a series of gentle head movement tasks. You will be asked to move your head from side to side or up and down in a manner which is comfortable to you.

6. Will my taking part in this study be kept confidential?

All information collected about you will be kept confidential and be used for research purposes only. It will be stored anonymously on computer for analysis. The information you give will not be available to anyone apart from the study investigator. Any information that is published will have your name removed so that you cannot be recognised from it.

7. What will happen to the results of the research?

The results of the study will form part of the study investigator's PhD theses. They may also be published. A copy of the results can be obtained from the main study investigator.

8. Who is organising and funding the research?

The study is being organised by the main study investigator who is a PhD candidate at the University of Southampton and an employee of the AECC. There is no direct funding of this project.

9. Who has reviewed the study?

The AECC Ethics Committee.

10. Who to contact for further information

The main study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

Thank you for participating in this study.

G. Rix BSc(Chiro), DC, FCC  
Main Study Investigator

This patient information sheet is for you to keep together with a copy of the consent form for you to sign

**Subject ID Number:**



## **Consent Form**

**Study Title:** Inter & Intra Examiner Reliability and Concurrent Validity of the Zebris CMS 70P and Laser Pointer Methods for Measuring Cervicocephalic Kinaesthetic Sensibility

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my legal rights being affected.
3. I agree to take part in the above study.

Name of subject

Date

Signature

Name of person taking consent

Date

Signature

The main study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

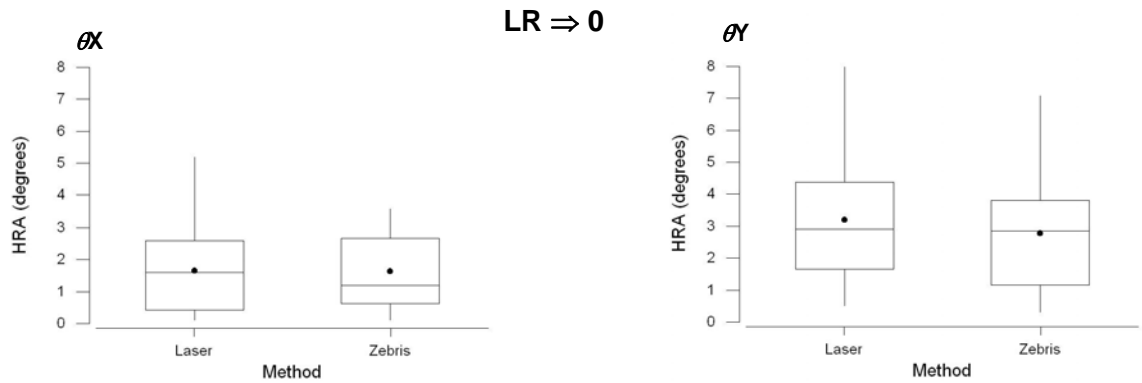
Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

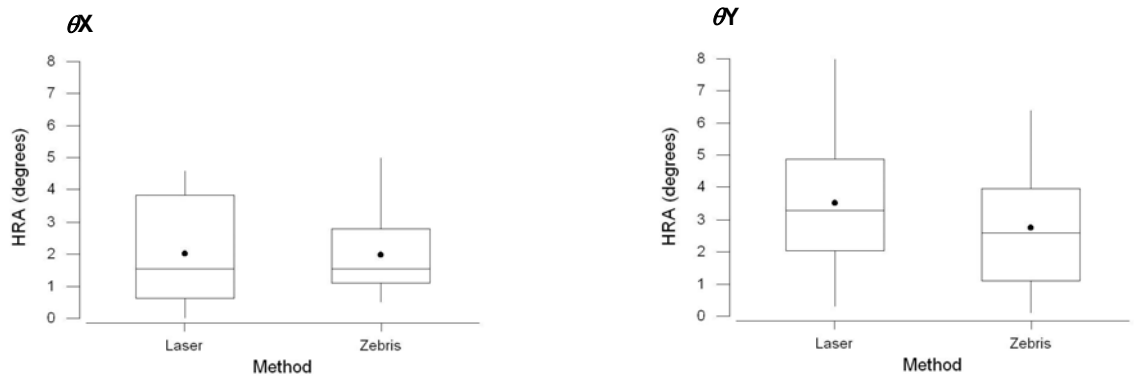
1 copy for subject; 1 copy for investigator

## Examiner 1

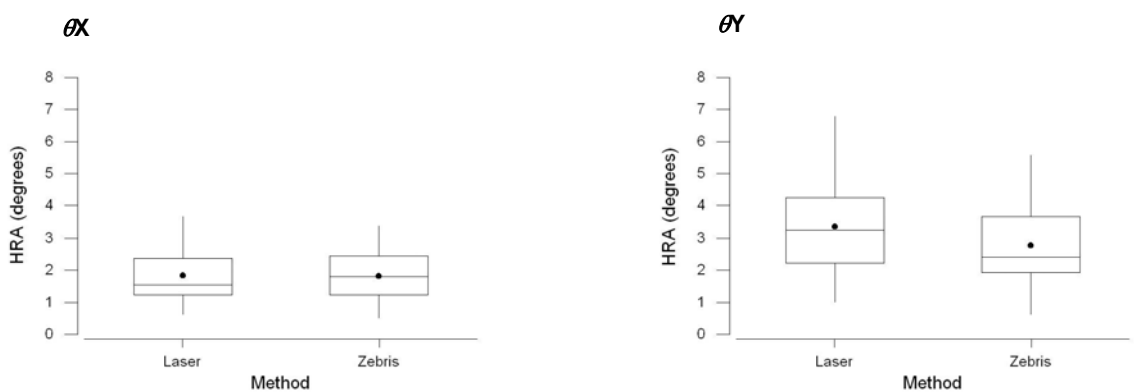
### Test



### Retest



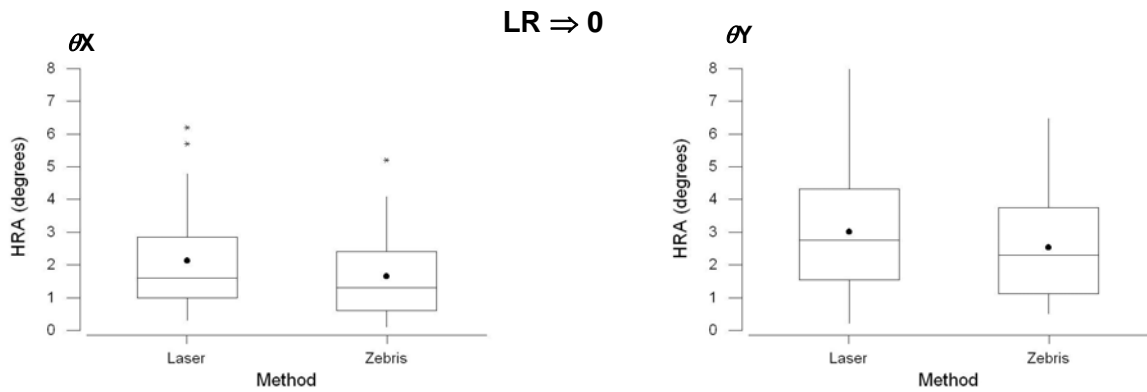
### Test +



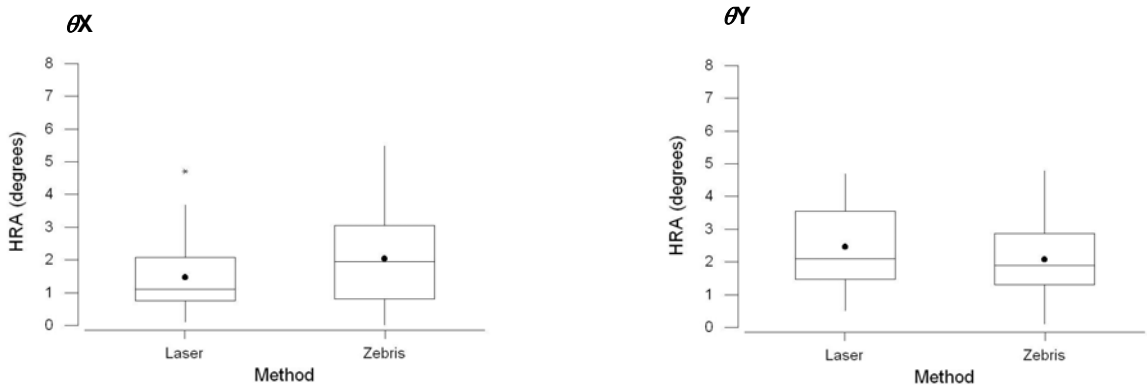
**Figure 64.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the  $LR \Rightarrow 0$  trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers.

## Examiner 2

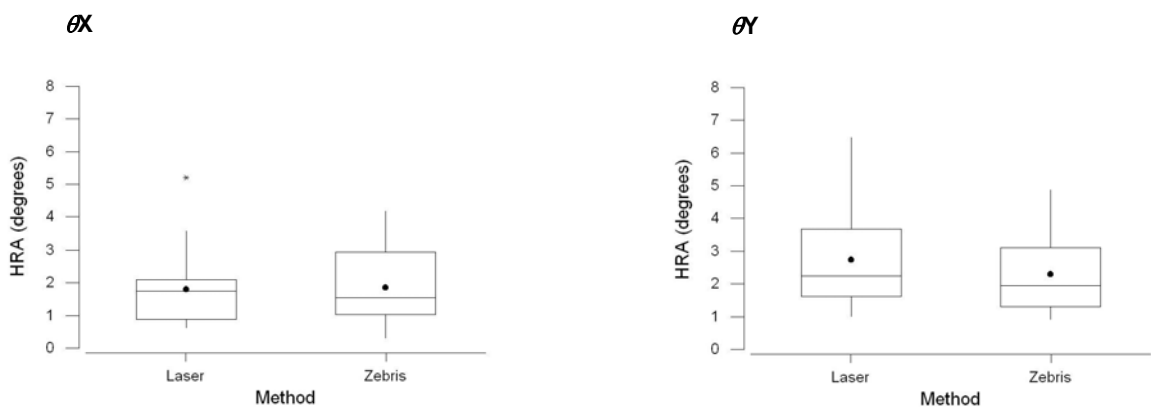
### Test



### Retest



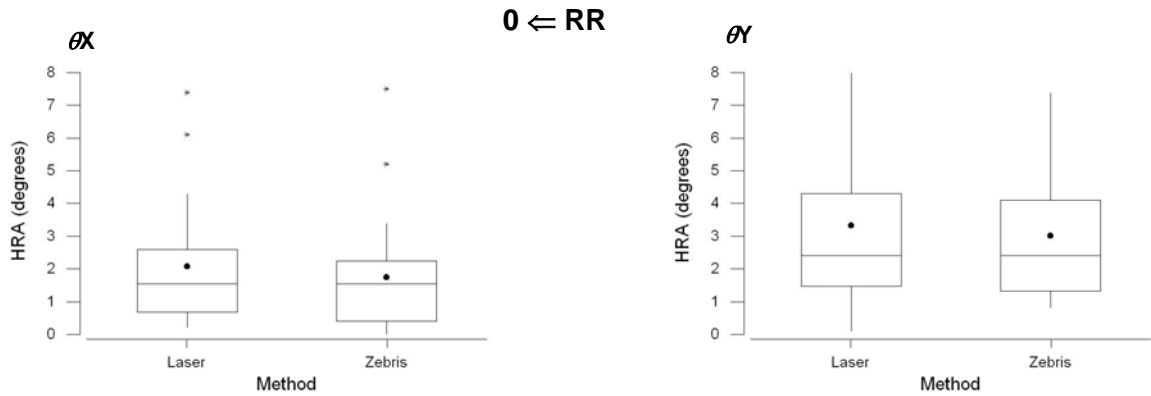
### Test + Retest



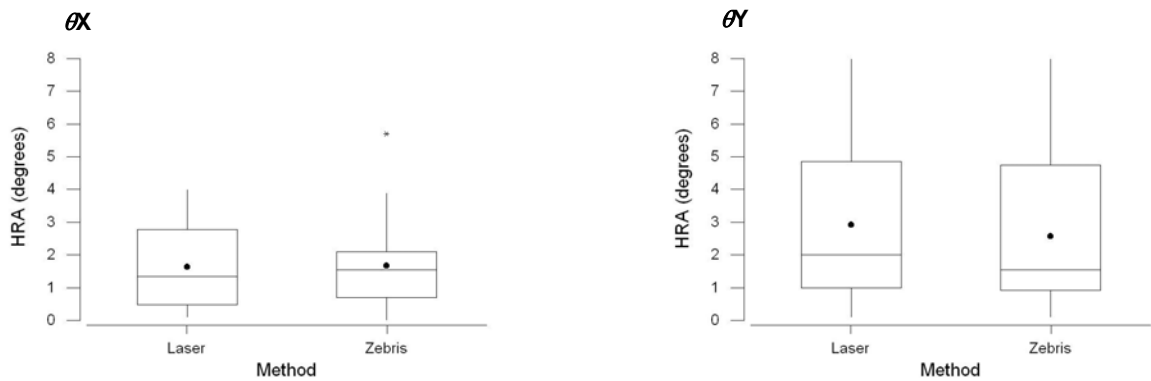
**Figure 65.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the  $LR \Rightarrow 0$  trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers.

## Examiner 1

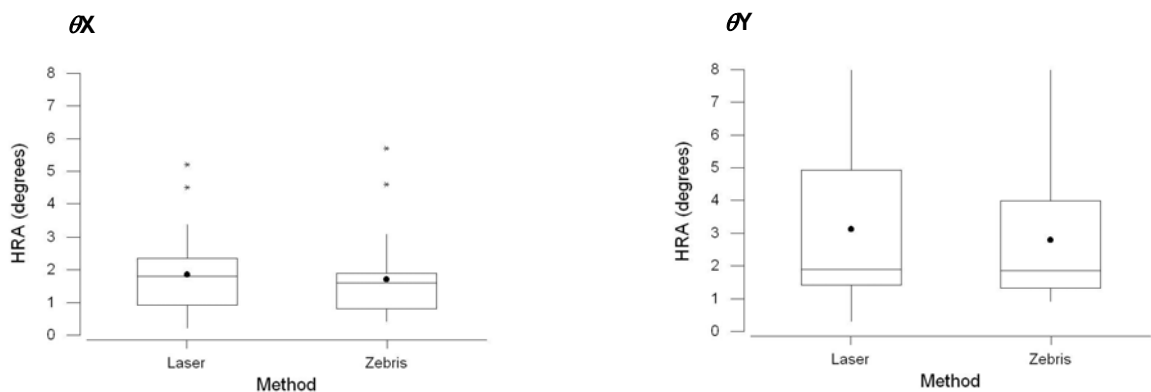
### Test



### Retest



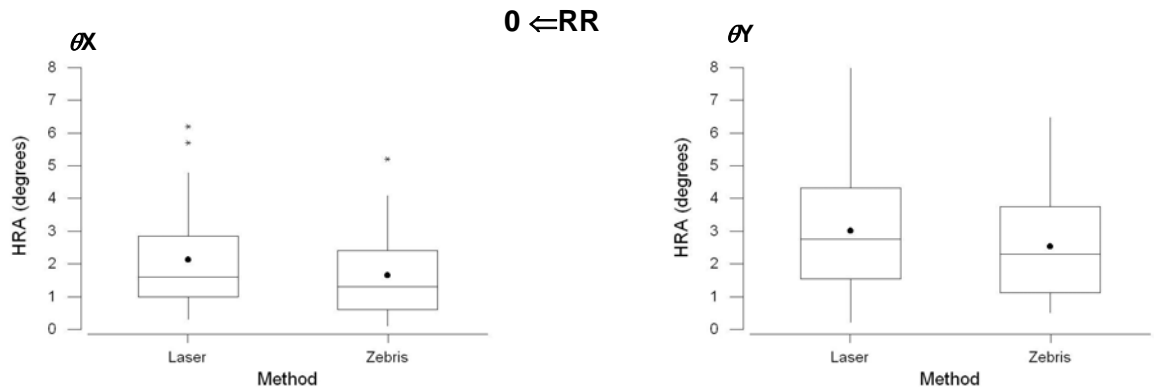
### Test + Retest



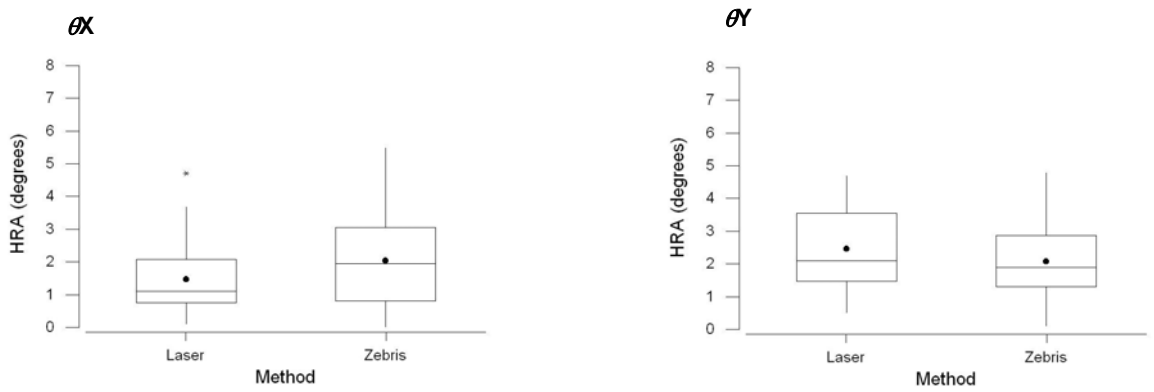
**Figure 66.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the  $RR \Rightarrow 0$  trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers.

## Examiner 2

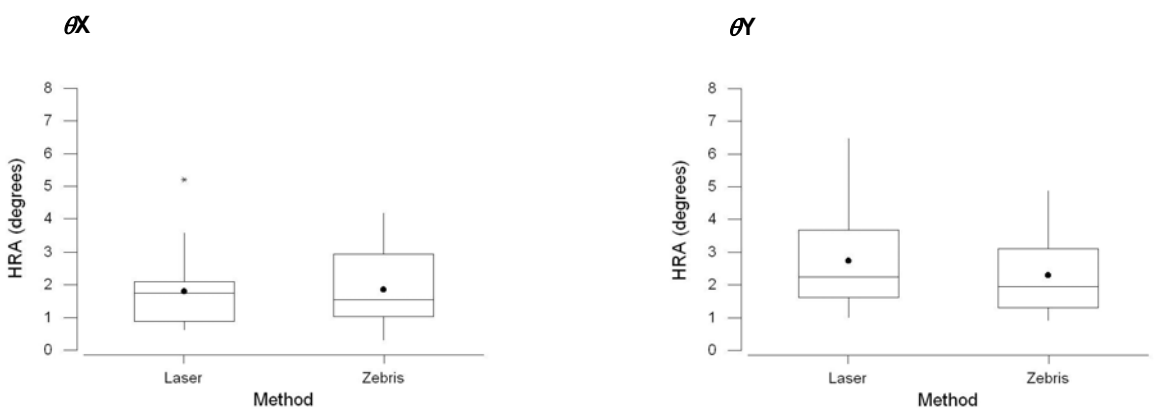
### Test



### Retest



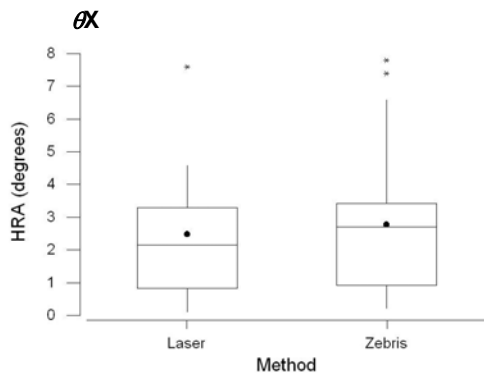
### Test + Retest



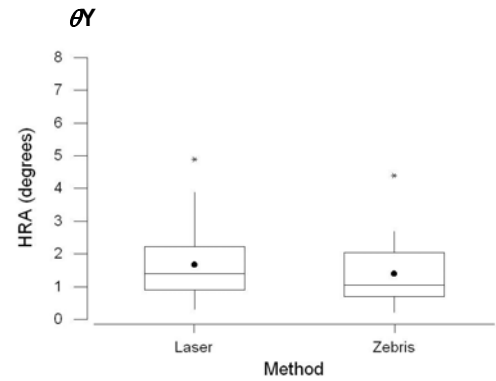
**Figure 67.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the  $RR \Rightarrow 0$  trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers.

## Examiner 1

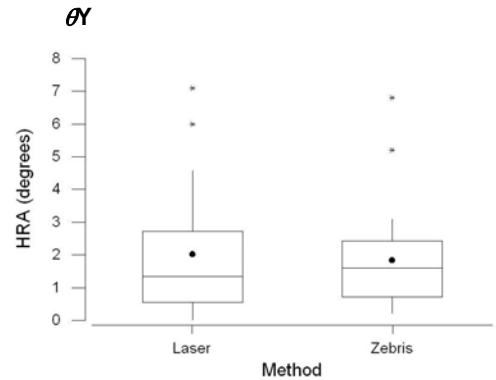
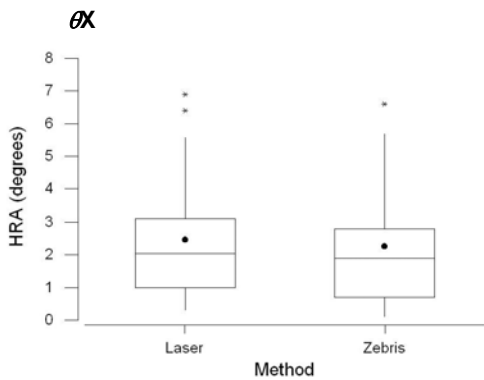
Test



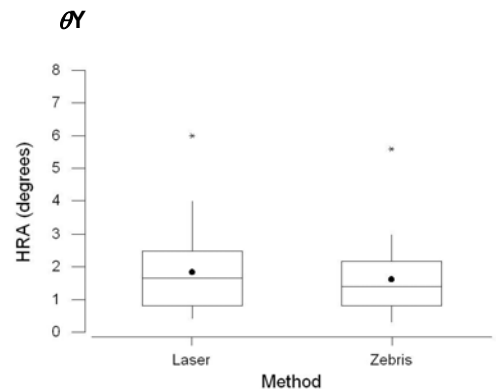
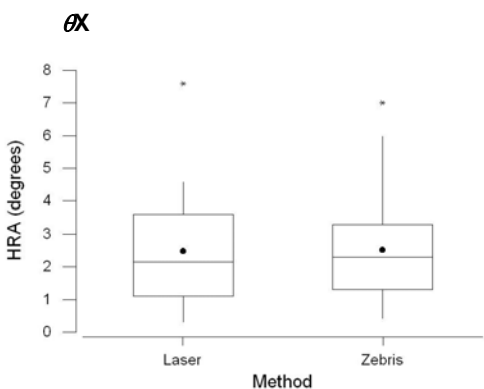
Ext  
↓  
0



Retest



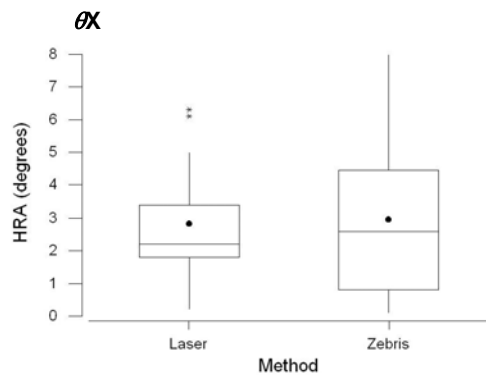
Test + Retest



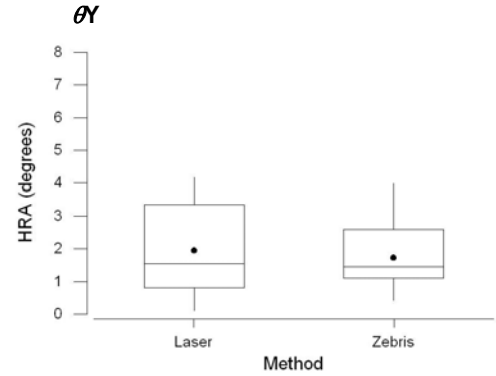
**Figure 68.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Ext  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers

## Examiner 2

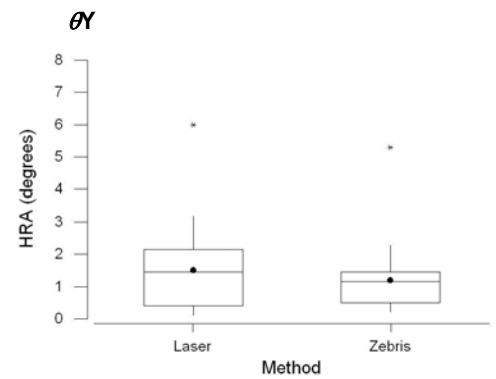
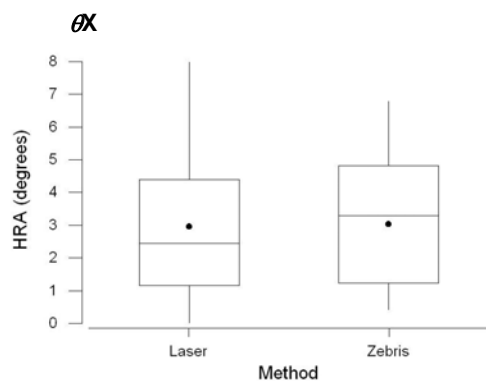
### Test



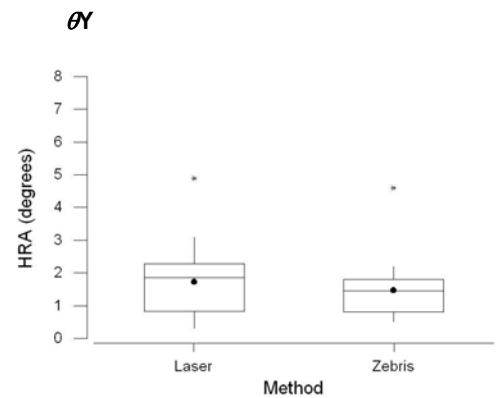
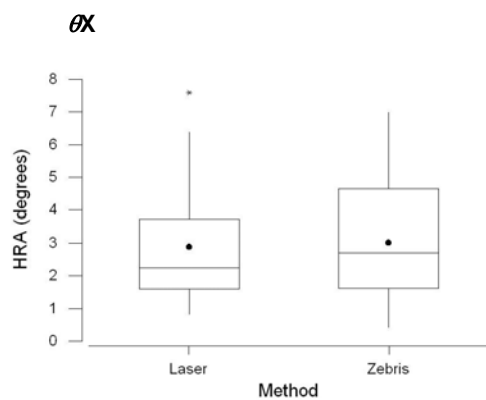
Ext  
↓  
0



### Retest



### Test + Retest

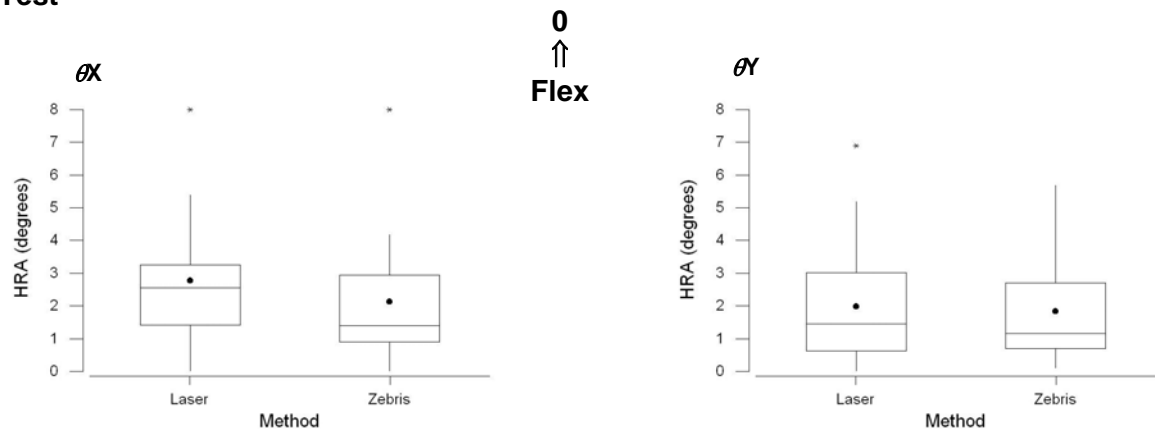


**Figure 69.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 2 for the Ext  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers

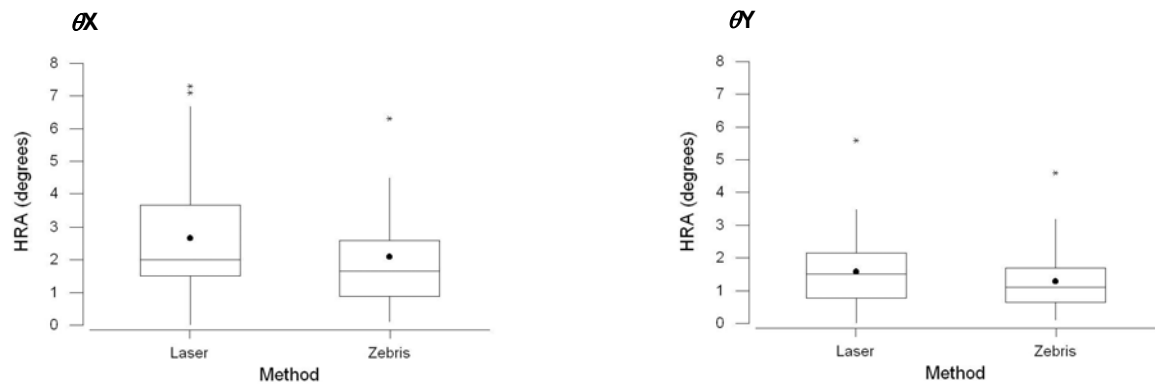


## Examiner 1

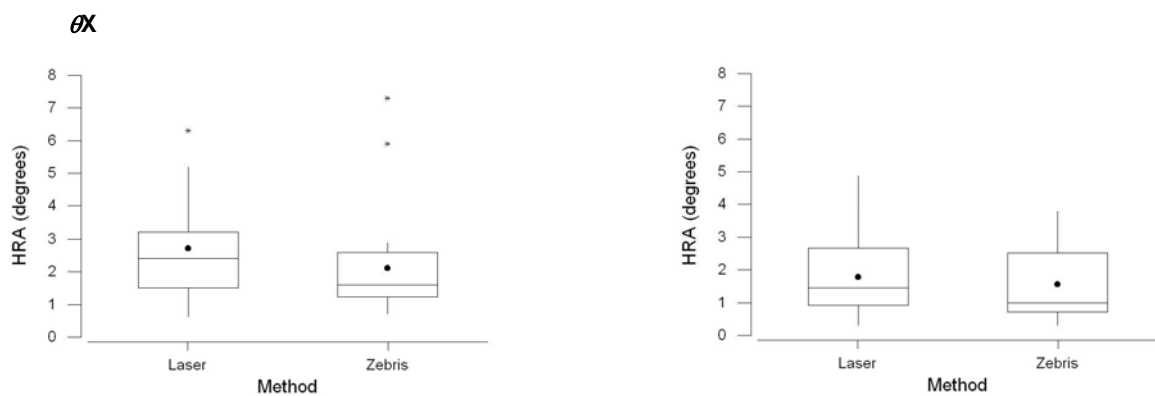
### Test



### Retest



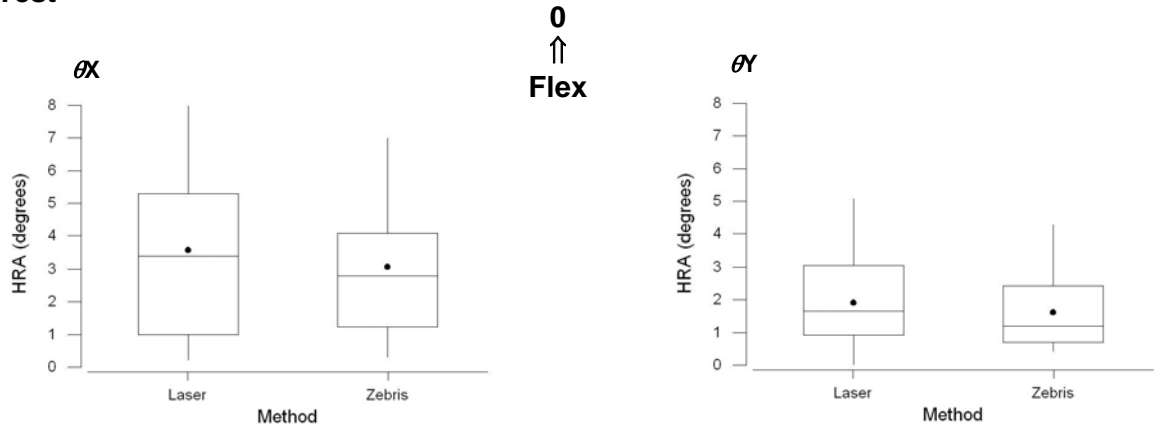
### Test + Retest



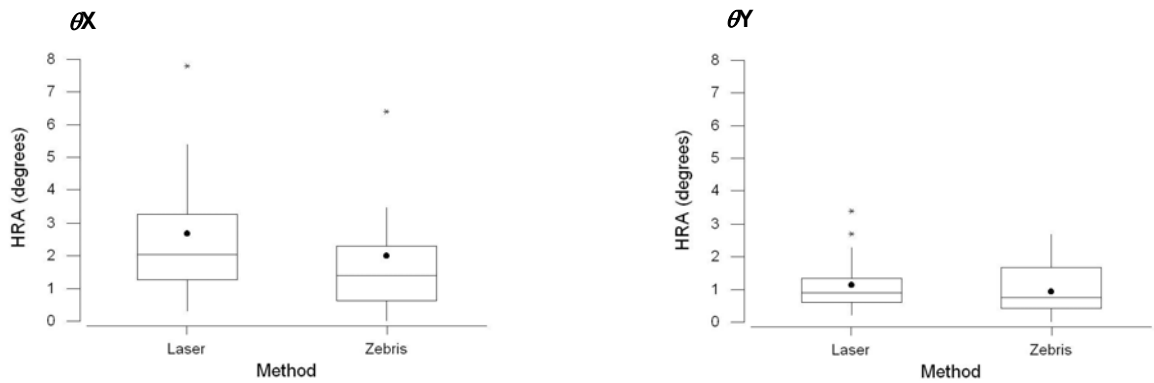
**Figure 70.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Flex  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5 (Q3 - Q1)$  Upper Limit:  $Q1 + 1.5 (Q3 - Q1)$ ) (Field, 2005). \* represent outliers

## Examiner 2

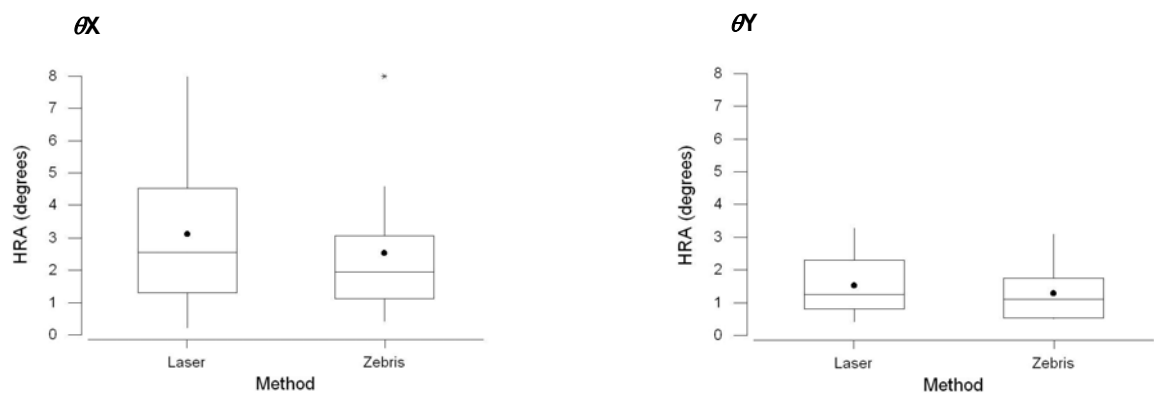
### Test



### Retest

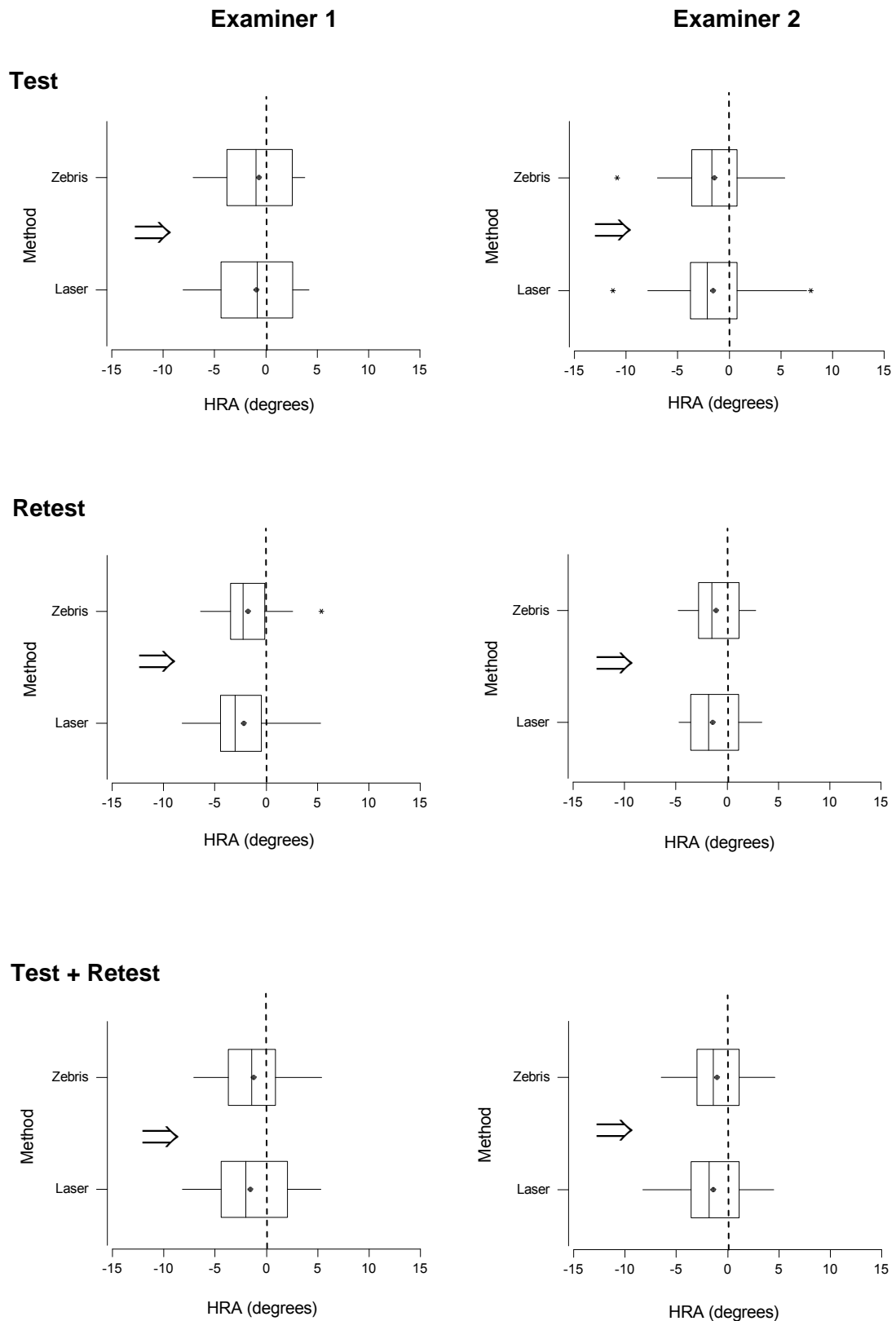


### Test + Retest



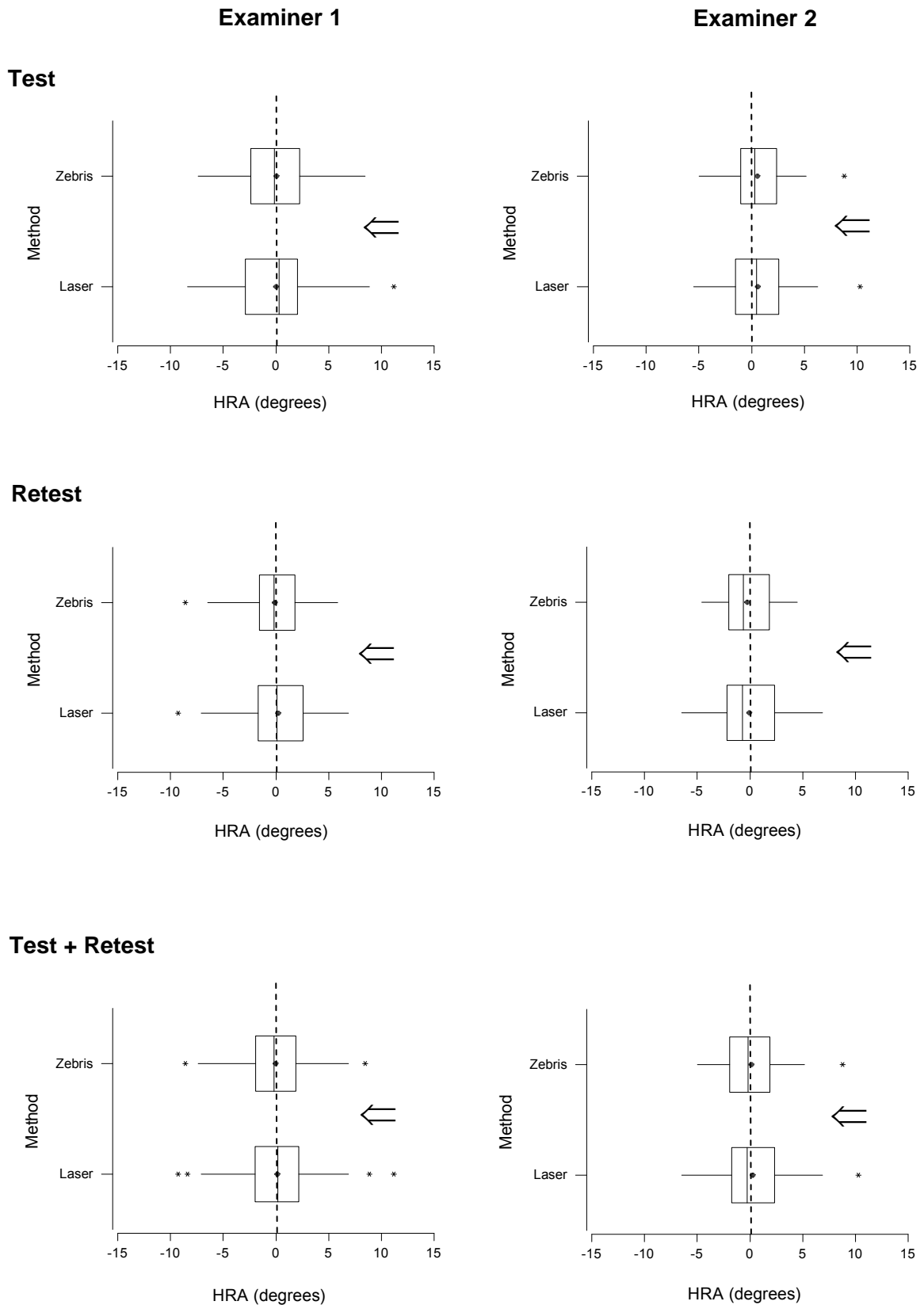
**Figure 71.** Box plots of absolute horizontal ( $\theta Y$ ) and vertical ( $\theta X$ ) repositioning error (degrees) with examiner 1 for the Flex  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5 (Q3 - Q1)$  Upper Limit:  $Q1 + 1.5 (Q3 - Q1)$ ) (Field, 2005). \* represent outliers

LR  $\Rightarrow$  0

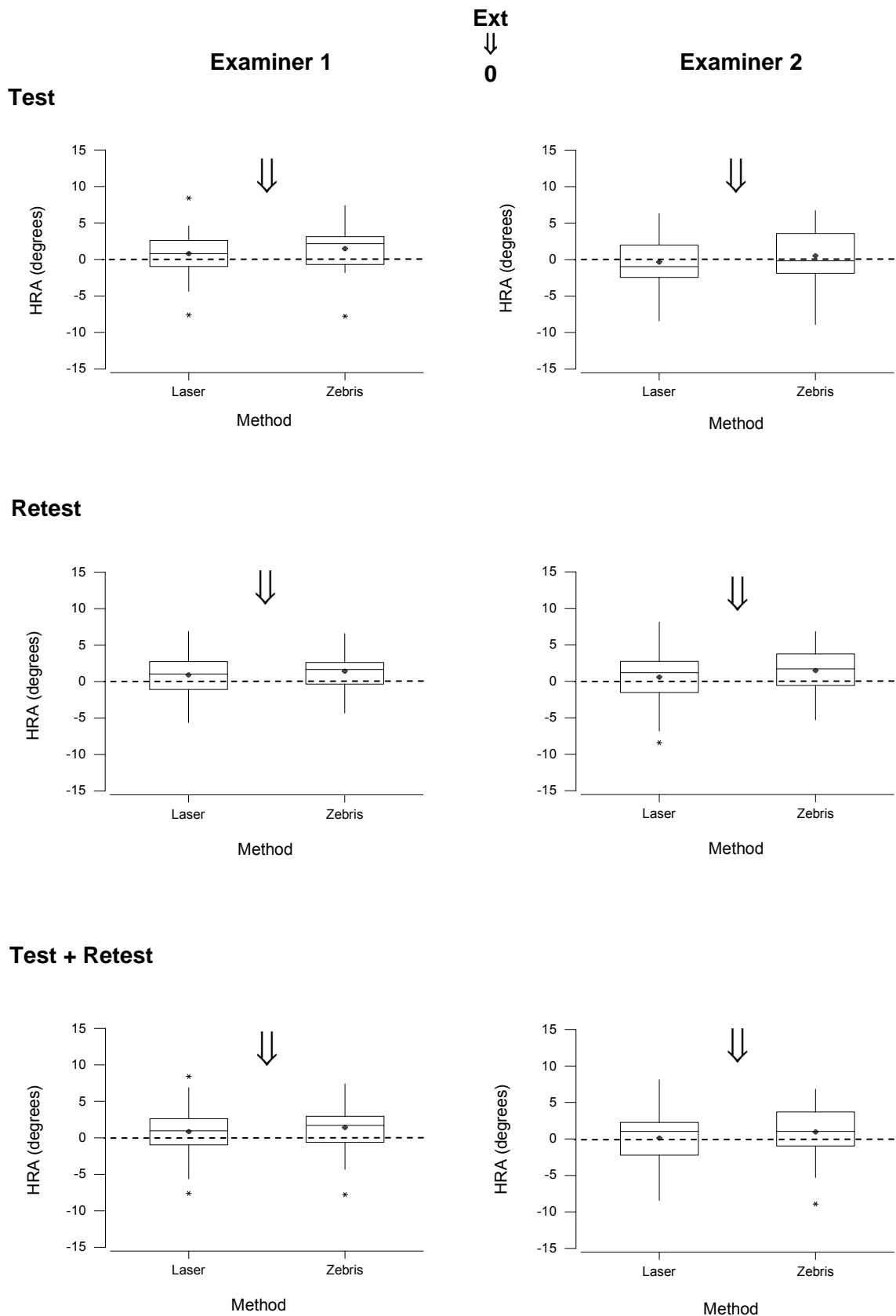


**Figure 72.** Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the LR  $\Rightarrow$  0 trial. The middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5 (Q3 - Q1)$  Upper Limit:  $Q1 + 1.5 (Q3 - Q1)$ ) (Field, 2005). \* represent outliers

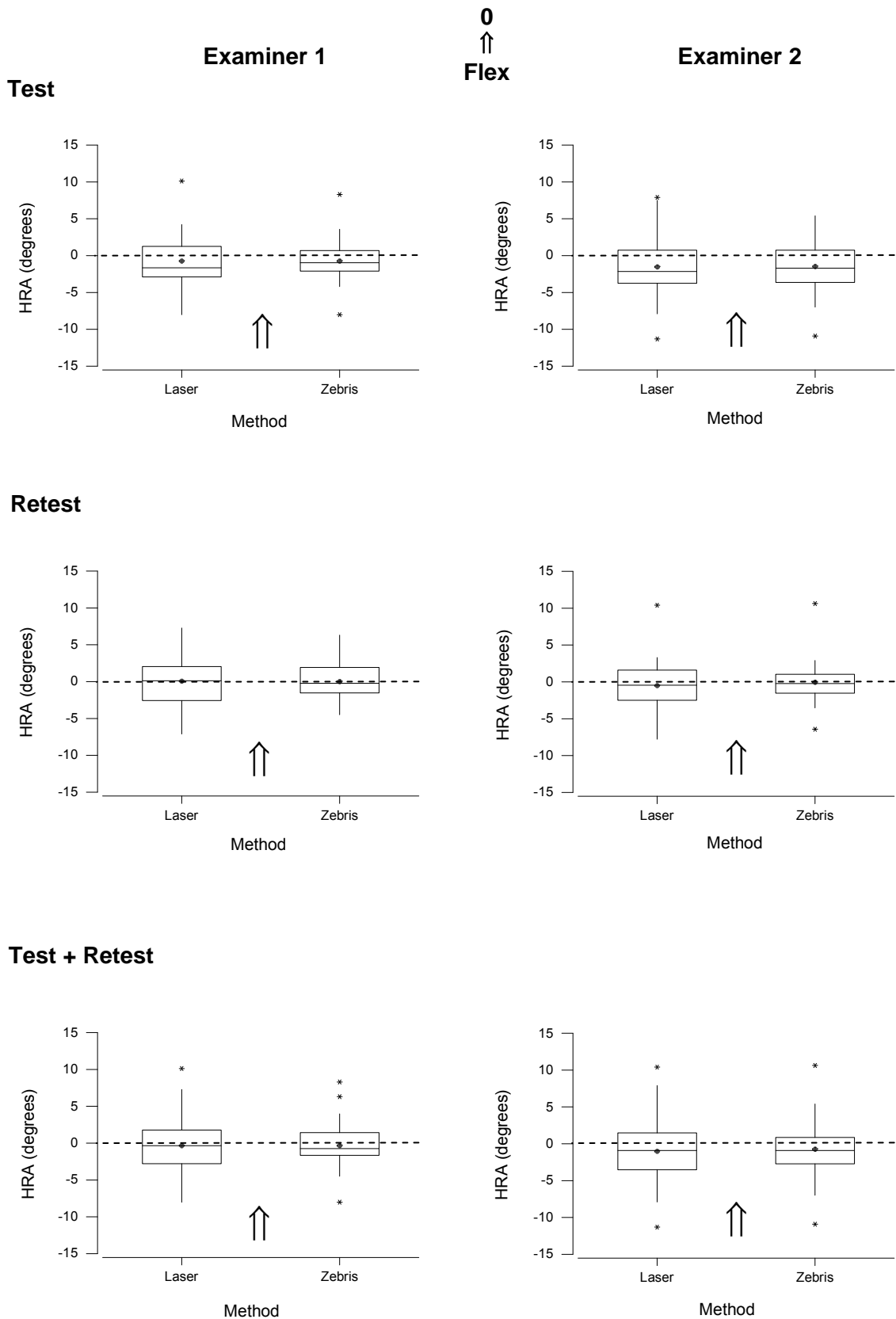
0 ← RR



**Figure 73.** Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta_Y$ ) with examiners 1 and 2 for the RR  $\Rightarrow$  0 trial. The middle vertical bar represents the median value; the box left and right sides represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5 (Q3 - Q1)$  Upper Limit:  $Q1 + 1.5 (Q3 - Q1)$ ) (Field, 2005). \* represent outliers



**Figure 74.** Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta_Y$ ) with examiners 1 and 2 for the Ext  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5(Q3 - Q1)$  Upper Limit:  $Q1 + 1.5(Q3 - Q1)$ ) (Field, 2005). \* represent outliers



**Figure 75.** Box plots of the signed repositioning error representing over/undershoot characteristics ( $\theta Y$ ) with examiners 1 and 2 for the Flex  $\Rightarrow$  0 trial. The middle horizontal bar represents the median value; the box bottom and top represent the interquartile range (25th and 75th percentile); the mean value is represented by the dot; the line whiskers represent the lowest and highest observations that are still inside the region defined by the following limits (Lower Limit:  $Q1 - 1.5 (Q3 - Q1)$  Upper Limit:  $Q1 + 1.5 (Q3 - Q1)$ ) (Field, 2005). \* represent outliers

## Appendix V

### Neck Pain Research Study

Study Investigator: Dr George D. W. Rix BSc (Chiro), DC, FCC

#### SCREENING QUESTIONNAIRE

For Office Use Only

File No:

1. What is this study?

The study will involve collecting information on the control of gentle voluntary head movements and the ability to determine what is vertical. It is hoped that this information will shed more light on the usefulness of these tests, the nature of neck dysfunction and inform the possible direction of future neck pain management strategies.

2. Who is conducting the study?

The study is being organised by the study investigator (see above & below) who is a PhD candidate at the University of Southampton and the Lead Lecturer for Neurology at the AECC.

3. What is this questionnaire?

This is an initial screening questionnaire to identify patients who may be eligible for the study. Completion of this form does NOT represent your agreement to participate. You will be given an information sheet before you decide whether or not to take part in the study. This will give you more details about what will happen to you if you agree to take part.

4. What will happen after completing the questionnaire?

The study investigator will review your clinical notes\* and if you are a suitable participant, he will contact you by phone to ask if you are willing and able to proceed to the next part of the study.

#### Please answer the questions below and return this form to clinic reception

<b>Q1</b>	Surname:	First Name:	
<b>Q2</b>	Age (years)		
<b>Q3</b>	Do you suffer from neck pain?	Yes	No
<b>Q4</b>	Is neck pain the main reason for you visiting us today?	Yes	No

#### Consent

1. I understand that sections of any of my clinical\* notes may be looked at by the investigator where it is relevant to my taking part in the study. I give permission for this individual to have access to my records.
2. I understand that I may be contacted by phone by the investigator regarding participation in this study. I give permission to be contacted by him for this purpose.

Signature of patient

Date

The study investigator: George D. W. Rix; Lead Tutor – Clinical Neurology, Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF. Tel: (01202) 436200 E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk). \*Chiropractor (AECC) held clinical notes

## Control Subject Information Sheet

### **Study Title: CERVICOCEPHALIC KINAESTHETIC SENSIBILITY AND SUBJECTIVE VISUAL VERTICAL PERCEPTION IN PATIENTS WITH CHRONIC, NON-TRAUMATIC CERVICAL SPINE PAIN**

Study Centre: Dept of Academic Affairs, Anglo-European College of Chiropractic (AECC), 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF. Tel: (01202) 436200

#### 11. Introduction

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and ask the investigator if there is anything that is not clear or if you would like more information.

Consumers for Ethics in Research (CERES) publish a leaflet entitled "Medical Research and You". This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained from CERES PO BOX 1365, London N10 0BW

#### 12. Purpose of the Study

The study will involve collecting information on the control of voluntary head movements and the ability to determine what is vertical. It is hoped that this information will shed more light on the usefulness of these tests described in section 5.

#### 13. Why have I been chosen?

You have been chosen because you do not suffer from neck pain and are otherwise healthy.

#### 14. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part you are still free to withdraw at anytime and without giving a reason.

**PLEASE TURN OVER**



15. What will happen to me if I take part?

You will be asked to attend the AECC outpatient clinic for a series of assessment procedures. These are as follows:

- Complete two further short questionnaires regarding your neck pain.
- Perform a series of gentle head movement tasks. You will be asked to move your head from side to side or up and down in a manner which is comfortable to you.
- Perform a series of computer based perception tasks. This will involve a series of tasks where you orient a line on a computer screen to a vertical position.

16. Will my taking part in this study be kept confidential?

All information collected about you will be kept confidential and be used for research purposes only. It will be stored anonymously on computer for analysis. The information you give will not be available to anyone apart from the study investigator. Any information that is published will have your name removed so that you cannot be recognised from it.

17. What will happen to the results of the research?

The results of the study will form part of the study investigator's PhD thesis. They may also be published. A copy of the results can be obtained from the main study investigator.

18. Who is organising and funding the research?

The study is being organised by the main study investigator who is a PhD candidate at the University of Southampton and an employee of the AECC. There is no direct funding of this project.

19. Who has reviewed the study?

The AECC Ethics Committee.

20. Who to contact for further information

The main study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

Thank you for participating in this study.

G. Rix BSc(Chiro), DC, FCC  
Main Study Investigator

This patient information sheet is for you to keep together with a copy of the consent form for you to sign

## Patient Information Sheet

### **Study Title: CERVICOCEPHALIC KINAESTHETIC SENSIBILITY IN PATIENTS WITH CHRONIC, NON-TRAUMATIC CERVICAL SPINE PAIN**

Study Centre: Dept of Academic Affairs, Anglo-European College of Chiropractic (AECC), 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF. Tel: (01202) 436200

#### 21. Introduction

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and ask the investigator if there is anything that is not clear or if you would like more information.

Consumers for Ethics in Research (CERES) publish a leaflet entitled "Medical Research and You". This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained from CERES PO BOX 1365, London N10 0BW

#### 22. Purpose of the Study

The study will involve collecting information on the control of voluntary head movements and the ability to determine what is vertical. It is hoped that this information will shed more light on the usefulness of these tests (section 5), the nature of neck dysfunction and inform the possible direction of future neck pain management strategies.

#### 23. Why have I been chosen?

You have been chosen because the neck complaint you have is of the correct type and duration for this particular study.

#### 24. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part you are still free to withdraw at anytime and without giving a reason. Participation or withdrawal from the study will not affect the treatment you receive.

**PLEASE TURN OVER**

25. What will happen to me if I take part?

Before your first chiropractic treatment visit, you will be asked to attend the AECC outpatient clinic for a series of assessment procedures. These are as follows:

- Complete two further short questionnaires regarding your neck pain.
- Perform a series of gentle head movement tasks. You will be asked to move your head from side to side or up and down in a manner which is comfortable to you.
- Perform a series of computer based perception tasks. This will involve a series of tasks where you orient a line on a computer screen to a vertical position.

26. Will my taking part in this study be kept confidential?

All information collected about you will be kept confidential and be used for research purposes only. It will be stored anonymously on computer for analysis. The information you give will not be available to anyone apart from the study investigator. Any information that is published will have your name removed so that you cannot be recognised from it.

27. What will happen to the results of the research?

The results of the study will form part of the study investigator's PhD thesis. They may also be published. A copy of the results can be obtained from the study investigator.

28. Who is organising and funding the research?

The study is being organised by the study investigator who is a PhD candidate at the University of Southampton and an employee of the AECC. There is no direct funding of this project.

29. Who has reviewed the study?

The AECC Ethics Committee.

30. Who to contact for further information

The study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

Thank you for participating in this study.

G. Rix BSc(Chiro), DC, FCC  
Study Investigator

This patient information sheet is for you to keep together with a copy of the consent form for you to sign

## Control Subject Consent Form

**Subject ID Number:**



### Consent Form

**Study Title: CERVICOCEPHALIC KINAESTHETIC SENSIBILITY IN PATIENTS WITH CHRONIC, NON-TRAUMATIC CERVICAL SPINE PAIN**

4. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.
5. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my legal rights being affected.
6. I agree to take part in the above study.

Name of subject

Date

Signature

Name of person taking consent

Date

Signature

The main study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

1 copy for subject; 1 copy for investigator

## Patient Consent Form

**Patient ID Number:**



### Consent Form II

**Study Title: CERVICOCEPHALIC KINAESTHETIC SENSIBILITY AND SUBJECTIVE VISUAL VERTICAL PERCEPTION IN PATIENTS WITH CHRONIC, NON-TRAUMATIC CERVICAL SPINE PAIN**

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without my chiropractic care or legal rights being affected.
3. I understand that sections of any of my clinical\* notes may be looked at by the study investigator where it is relevant to my taking part in the study. I give permission for this individual to have access to my records.
4. I agree to take part in the above study.

Name of patient

Date

Signature

Name of person taking consent

Date

Signature

The study investigator: George D. W. Rix; Dept of Academic Affairs, Anglo-European College of Chiropractic, 13-15 Parkwood Road, Bournemouth, Dorset BH5 2DF.

Tel: (01202) 436200

E-mail: [grix@aecc.ac.uk](mailto:grix@aecc.ac.uk).

1 copy for patient; 1 copy for investigator; 1 copy to be kept with clinical notes\*

\*Chiropractor (AECC) held clinical notes

## AECC Neck Pain Research Project

This questionnaire is about your neck pain. The information will be treated anonymously and in confidence  
Please fill in the relevant boxes with crosses (NOT TICKS)

**Q1.** How long have you suffered from **THIS** episode of neck pain (eg., 2 days, 7 weeks, 3 months, 2 years etc)? \_\_\_\_\_

**Q2.** Has your neck pain been continuous (continuous pain every day) since **THIS** episode started?

- Yes
- No →

**Q3.** On average, how often has your neck pain occurred since **THIS** episode started?

- More than once a day     once a day     once every two days
- once every three days     once a week     Other (please state below)
- \_\_\_\_\_

**Q4.** On average, how long has each of the neck pain 'attacks' lasted ?

- less than one hour     1-2 hours     2-4 hours
- Other (please state) \_\_\_\_\_

**Q5.** Is **THIS** episode the result of a specific injury/trauma?

- Yes →
- No

**Q6.** What was the nature of the injury/trauma?

- Motor vehicle accident/collision     Other (please state) \_\_\_\_\_
- \_\_\_\_\_

**Q7.** How soon after the injury/trauma did you feel the pain?

- less than 48 hours after     more than 48 hours after

**Q8.** Have you suffered from previous episodes of neck pain prior to this current episode?

- Yes →
- No

**Q9.** When was your **FIRST** episode of neck pain (eg., 6 months ago, 2 yrs ago)? \_\_\_\_\_

**Q10.** Was this **FIRST** episode the result of a specific injury/trauma?

- Yes
- No

**Q11.** What was the nature of the injury/trauma?

- Motor vehicle accident/collision
- Other (please state) \_\_\_\_\_
- \_\_\_\_\_

**Q12.** How soon after the injury/trauma did you feel the pain?

- less than 48 hours after     more than 48 hours after

**Q14.** Since your **FIRST** episode of neck pain, how many motor vehicle accident/collisions have you had?

**Q15.** How many episodes of neck pain have you suffered since your **FIRST** episode?

- One     Two     Three
- Four     Five     Other (please state) \_\_\_\_\_

**Q16.** On average, how long has each of these episodes neck pain lasted ?

- less than one week     1-2 weeks     2-4 weeks
- Other (please state) \_\_\_\_\_

### For Office Use Only

Subject No:  
Date: