Clinical Management and Measurement of Upper Limb Disability in Neck Pain Patients

Ahmad Salah Eldin Alreni

A thesis submitted in partial fulfilment of the requirement of Sheffield Hallam University for the degree of Doctor of Philosophy

July 2018

Abstract

There is a strong relationship between non-specific neck pain (NS-NP) and upper limb disability (ULD). Optimal management of NS-NP should incorporate upper limb (UL) rehabilitation and therefore include evaluation of ULD using suitable UL outcome measure (OM) in the assessment and during the management process. However, there is no clear guidance regarding the suitability of available measures alongside a lack of information on how physiotherapists in the United Kingdom (UK) measure and rehabilitate their patients with NS-NP. The purpose of this thesis was to explore the clinical measurement and management of ULD in patients with NS-NP.

The quantitative research approach adopted by this thesis enabled the researcher to gain a deeper understanding of the clinical measurement and in turn rehabilitation of ULD in patients with NS-NP, and build on knowledge acquired throughout the period of study. In order to support this methodology, a positivist philosophical stance was adopted.

A systematic review was completed to identify all available UL OMs that were used for patients with neck pain (NP) and to make recommendations about those that are suitable for use in clinical practice and research. A survey with a national sample of physiotherapists was completed to establish current physiotherapeutic management of NS-NP and ULD in the UK. This was followed by a validation study aimed at exploring the acceptability and feasibility of the Single Arm Military Press (SAMP) test. Subsequently, a second validation study was completed to explore the reliability and validity of the SAMP test in female patients with NS-NP and healthy subjects.

The systematic review identified five measures but quality issues prevented a clear recommendation for any of the identified instruments. The survey highlighted substantial gaps in current evidence-based practice of UK physiotherapists regarding the measurement of patients with NS-NP and associated deficits in the measurement and management of ULD in this population. Subsequently, a validation study established the acceptability and feasibility of the SAMP test using a 1-kg hand weight in female patients with NS-NP. In the second validation study, the SAMP test was found to be a reliable and valid UL instrument for female patients with NS-NP.

This thesis provided preliminary evidence that the SAMP test is an acceptable, feasible, valid and reliable measure of ULD for female patients with NS-NP and of its suitability for use in clinical practice and research. The SAMP test can be used by clinicians to improve their assessment of UL functional capacity and to suggest management strategies for patients with NS-NP. Further longitudinal studies are required to evaluate the further validity and reliability of the SAMP test in older and younger female patients, and male patients using additional examiners and additional populations. Further studies are required to establish the responsiveness of the SAMP test in patient populations with all types of NP.

Table of contents

Chapter 7: Summary, discussion and conclusion

References

List of appendices

List of figures

List of tables

List of abbreviations

AH = Asmaa Hamdy

ASEA = Ahmad Salah Eldin Alreni

AL = Anna Lowe

ANOVA = Analysis of Variance

APTA = American Physical Therapists Association

AUC = Area Under Curve

BH = Basma Hassan

BQ = Bournemouth Questionnaire

CI = Confidence Interval

COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments

CSP = Chartered Society of Physiotherapy

CSOQ = Cervical Spine Outcome Questionnaire

DASH = Disability of Arm, Shoulder and Hand Questionnaire

DC = Galvanic Current

ENS = Galvanic Current, Electrical Nerve Stimulation

ES = Effect Size

FIT-HaNSA = Functional Impairment Test-Hand, and Neck/Shoulder/Arm

GDP = Gross Domestic Product

GE = Ghada Eid

HA = Heba Abdo

ICC = Interclass Correlation Coefficient

IRT = Item Response Theory

ISOQOL = International Society for Quality Of Life research

LoA = Limits of Agreement

MCS 36/12 = Mental Component Summary Scale 36/12

MDC = Smallest Detectable Change

MDT = The Mechanical Diagnostic and Therapy

MIC = Minimal Important Change

MSCP = Musculoskeletal Association of Chartered Physiotherapists

NDI = Neck and Disability Index

NHS = National Health Services

NHP = Nottingham Health Profile

NP = Neck Pain

NPQ = Northwick Park Neck Pain Questionnaire

NRS = Numeric Rating Scale

NS-NP = Non-Specific Neck Pain

NULI = Neck and Upper Limb Index

OMs = Outcome Measures

PBOMs = Performance-Based Outcome Measures

PCS = Pain Catastrophizing Scale

PCS 36/12 = Physical Component Summary Scale 36/12

PEMF = Pulsed Electromagnetic Fields Therapy

PFActS-C = Pictorial Fear of Activity Scale–Cervical

PGT = Postgraduate Training

PPA = Physiotherapy Pain Association

PROMs = Patient-Reported Outcome Measures

PRISMA = The Preferred Reporting Items for Systematic Reviews and Meta-Analysis

PSE = Pain Self-Efficacy

PSFS = Patient Specific Functional Scale

QST = Quantitative Sensory Tests

ROC = Receiver Operator Curve

rMS = Repetitive Magnetic Stimulation

RULA = Rapid Upper Limb Assessment

RQ = Research Question

SAMP = Single Arm Military Press

SD = Standard Deviation

SDC = Smallest Detectable Change

SEM = Standard Error of Measurement

SFA = Shoulder Functional Assessment

SF-36/12 = Short Form 36/12 Questionnaire

Sim II = Baltimore Therapeutic Equipment Work Simulator II

SMc = Sionnadh McLean

SPSS = Statistical Package for Social Science

SRM = Standardised Response Means

TENS = Transcutaneous Electrical Nerve Stimulation

TP = Tanzila Potia

UK = United Kingdom

ULD = Upper Limb Disability

UL = Upper Limb

USA = United State of America

VAS = Visual Analogue Scale

WHO = World Health Organisation

σ 2 = Variance

Acknowledgements

First and foremost, I would like to thank Allah Almighty, the Most Gracious and the Most Merciful for giving me the strength, knowledge, ability and opportunity to undertake this research project and to complete it satisfactorily. Without Allah's blessing, this achievement would not have been possible.

In my journey towards this degree I have found a teacher, a friend, an inspiration and a pillar of support in my guide, Dr Sionnadh McLean (Director of Studies). She has been providing her wholehearted support and guidance throughout this research project. Her invaluable help and constructive feedback throughout the experimental and thesis work have contributed to the success of this research project. Without her able guidance, this thesis would not have been possible and I shall eternally be grateful to her for her assistance.

I would like to express my sincere appreciation and thanks to my supervisors, Dr Anil Gumber, Dr Karen Kilner and Deborah Harrop (Information Scientist) for their support and guidance throughout this research.

I would like to thank Anna Lowe and Tanzila Potia for their assistant in the screening process and quality assessment of studies in the systematic review

I would like to thank Cliona McRobert for the support in developing the Social Media Strategy in the survey as well as Sean Demack for the assistant with the survey.

I would like to thank Professor Ali Eldeeb and all the staff in the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital for their constant help and support throughout the recruitment and examination of patients in the validation studies.

I would also like to thank all UK-based physiotherapists who volunteered to participate in the UK national survey of neck pain as well as all patients and healthy subjects who also volunteered to participate in the validation studies.

Last but not the least, I would like to express my deepest gratitude to my family without whose support it would have been difficult to complete this research project. It is to my family and children I dedicate this work.

Chapter 1: Overview of the thesis

1.1 Introduction

The overall aim of this programme of research was to investigate the clinical measurement and management of ULD in adult, female patient populations with NS-NP. This chapter provides an introduction and a justification for undertaking the research, including its overall aims, philosophical framework, ethics, structure and an overview of the studies that were conducted as part of the research. A summary of the thesis chapters and a list of published and proposed peer reviewed publications and conference presentations relating to the PhD are provided.

1.2 Neck pain (NP)

In the 21st century, NP is a common musculoskeletal condition that causes substantial pain and disability. The Bone and Joint Decade 2000 - 2010 Task Force on NP and its associated disorders systematically reviewed the published literature on NP between 1980 - 2006 to produce a best evidence synthesis on the burden and determinants of NP (Hogg-Jonson et al. 2008). Evidence from the 101 included studies identified that the NP incidence rates range between 15.5 and 213 per 1000 - person years, while the 1-year prevalence of pain rates range around 30 - 50% with 1.7 - 11.5% of people experiencing disability because of their NP annually. NP was found to have higher prevalence for females than males with ratio ranging from 3.4:1.1. Furthermore, females demonstrated higher rates of visits to healthcare centres seeking treatment for their NP (males: 2.6 visits per 1000, 95% CI, 2.1 - 3.0; females: 3.5 visits per 1000, 95% CI, 3.0 – 4.0).

In another large and well-designed study, Hoy and colleagues (2010) collected data relating to the incidence, remission and prevalence of NP from all published and unpublished population-based studies conducted between 1980 - 2009 inclusive, which was for the purpose of assessing the global burden of NP throughout the world (Hoy et al. 2010). Surveys that had mainly focussed on general population with mild, moderate or severe NP were used to provide data with no language, age, gender or setting restriction. The findings provide evidence that the 1-year incidence of NP ranged from 10.4% to 21.3%, while the remission of NP, which was defined in the study as the rate at which NP has been completely resolved, at 1-year ranged from 33% to 65%. The overall prevalence of NP in the general population ranged between 0.4% and 86% (mean: 23.1%), point prevalence ranged from 0.4% to 41.5% (mean: 14.4%), and the 1-

year prevalence ranged from 4.8% to 79.5% (mean: 25.8%). This study also provided evidence that the incidence of NP is higher among females with an increased risk of developing NP until the 35-49-year age group alongside those with history of low back pain, poor psychological status, low job satisfaction, sedentary work posture, poor physical work environment and smoking. Subsequently, Hoy and colleagues (2014) collected data to estimate the global burden of NP in relation to its associated disability (Hoy et al. 2014). The prevalence of NP in 2010 was estimated to be 4.9% (95% CI: 4.6 to 5.3) and the disability YLDs (years of life lived with disability) had increased from 23.9 million (95% CI: 16.5 to 33.1) in 1990 to 33.6 million (95% CI: 23.5 to 46.5) in 2010. Disability because of NP was found to be higher in female patients (mean: 5.8%, 95% CI: 5.3 to 6.4) than in male patients (mean: 4.0%, 95% CI: 3.7 to 4.4) and the prevalence peaked at 45 years of age. This study reported that out of 291 musculoskeletal conditions, NP was ranked 4th highest in terms of disability when measured by YLDs and 21st in terms of overall burden.

In their epidemiological study, Thomas and colleagues (2004) collected data relating to the presence of pain and pain interference (disability) in older age people (50+ years) from three primary care general practice in the North Staffordshire (UK). Postal questionnaires were mailed to 11230 patients, of whom 7878 provided data regarding any pain with adjusted response rate of 71.3%. Of those providing data 22.8% in the age group (50-59), 22.9% in the age group (60-69), 17.7% in the age group (70-79) and 14.9% in the age group (80+) experienced NP which limited their daily activity (disability) at some point during the previous month. The study reported that pain and its associated disability were higher in female older patients than males. In another recent study, Scarabottolo and colleagues (2017) conducted a large-scale trial, in which 1011 adolescents between 10-17 years of age completed questionnaires relating to their back and neck pain. Of those who completed the questionnaires, 17.4% experienced NP at some point during the previous week. The prevalence estimates of NP were higher in older adolescents compared to younger adolescents but reasonably similar across gender.

There is strong epidemiological evidence that NP is also a common and disabling musculoskeletal condition in the UK. NP affects approximately 31% of the UK adult population at any one time with the majority experiencing recurrent or chronic

symptoms and 7.5 to 14% of those patients appear to experience some degree of disability because of their NP (Croft et al. 2001, Webb et al. 2003).

Whilst no studies have specifically explored the financial burden of NP and its associated disability in the UK, Borghouts and colleagues (1999) reported that the total annual cost of NP in the Netherland estimated US$ 686.2 million, which represent approximately 1% of the healthcare budget and 0.1% of the total GDP (Gross Domestic Product) in the Netherland (Borghouts et al.1999). Since both the Netherland and the UK are developed Western European countries, it would be reasonably acceptable to suggest that a similar proportion of healthcare and societal expenditure may be attributed to NP in the UK. These findings indicate that it is apparent that NP results in significant healthcare cost, work absenteeism and loss of productive capacity; NP is therefore a substantial socioeconomic burden for patients, employers, insurers and society.

For a majority of patients with NP, a pathoanatomical cause cannot be identified (Hoving et al. 2002, Walker-Bone et al. 2003, Binder 2007). Consequently, a wide variety of classification approaches has emerged. One such example of a classification system is patients with (1) serious spinal pathology requiring urgent medical attention, (2) neurological involvement and (3) non-specific neck pain (NS-NP) (Moffett and McLean 2006). NS-NP, which comprises approximately 80% of all NP patients, will be the focus of this thesis and is defined as “pain perceived as arising from anywhere within the region bounded superiorly by the superior nuchal line, inferiorly by the transvers line through the tip of the first thoracic spinous process, and laterally by the sagittal plans tangential to the lateral border of the neck” (Merskey and Bogduk 1994). This pain is not caused by any serious acute trauma, systematic disease, neurological disorder, or inflammatory conditions (Huisstede et al. 2007). Given that NP is more common in females (Thomas et al. 2004, Hog-Jonson et al. 2008, Haldeman et al. 2010, Hoy et al. 2010, Hoy et al. 2014) and female patients tend to have higher rate of seeking physiotherapy treatment for their NP (Freburger et al. 2005, Hog-Jonson et al. 2008), much of the research in this thesis focussed on female patients.

1.3 Upper limb disability (ULD) and neck pain (NP)

UL dysfunction is a common musculoskeletal condition (Walker-Bone et al. 2002). The prevalence of UL dysfunction at any given point of time has been estimated as 20% and

50% in the working population of Western industrial countries and the lifetime prevalence of UL dysfunction is greater than 70% (Walker-Bone et al. 2004, Huisstede et al. 2006). ULD can arise from a spectrum of clinical conditions, including NP (Huisstede et al. 2009). An extreme example of this is cervical radiculopathy which can lead to pain, motor weakness, sensory deficit and loss of function in the neck, shoulder, upper arm or forearm (Polstone 2007, Rhee et al. 2007). NS-NP has also been shown to have a considerable impact on UL function (Frank et al. 2005). In 2007, McLean and colleagues investigated the relationship between NS-NP and ULD in 151 patients with NS-NP who were recruited from four National Health Service (NHS) physiotherapy departments in the UK (McLean et al. 2007). A positive correlation was observed between the score of the Northwick Park Neck Pain Questionnaire (NPQ) that was used to measure the baseline of the NS-NP and the score of the Disability of Arm, Shoulder and Hand Questionnaire (DASH) that was used to measure the baseline of the ULD (Pearson's $r=0.799$, $p<0.001$, $n=142$). Furthermore, Stepwise Linear Regression analysis revealed that after adjusting for a range of other potential confounding variables, higher NPQ score ($B = 0.743$) and lower pain self-efficacy ($B = - 0.489$) predicted increased severity of the ULD ($R^2=0.713$; $p<0.001$, $n=100$). Subsequently, Osborn and Jull (2013) conducted a cross-sectional survey in adult Australian patients presenting for physiotherapy rehabilitation in the general community (n-103) which explored the proportion of NS-NP patients who experienced ULD and the nature of those UL activity. Moderate-high correlation were observed between the NDI score and the DASH score ($p = 0.669$; $p < 0.001$). In both studies approximately 80% of the NS-NP patients reported ULD in relation to activity that involve loading the UL such as lifting and repetitive overhead movement. These findings provide evidence of a strong relationship between NS-NP and ULD and that patients with the most severe NS-NP report the greatest severity of ULD. Both studies also recommended additional evaluation of UL functional capacity using suitable UL OM in the assessment and during the management of patients with NS-NP.

The mechanisms which cause NS-NP and ULD to co-exist are not clear but may relate to the mechanical attachment between the neck and the UL via skeletal, muscular and neural structures (McLean et al. 2011). For example, mechanical loading or repetitive movement of the UL may increase the mechanical load to the articular and ligamentous structures of the neck which may in turn provoke NP or create protective neck muscle spasms (Gorski and Schwartz 2003). Another possible mechanism is that patients with

NP may limit the functional use of their ULs because of neck provocation or poor pain self-efficacy (McLean et al. 2007, McLean et al. 2011). Consequently, a deconditioning effect may occur leading to a reduction in cardiovascular capacity and reduced strength and endurance in the neck/UL muscles, and this may lead to compensatory activity and excessive loading on the cervical structures (Smeets et al. 2006). Further investigation of these causal relationships is required, but to do this valid and reliable measures of ULD in patients with NP are required.

Clinical textbooks on the examination of patients with NS-NP often recommend simple screening of shoulder range of motion in order to rule in/out the presence of shoulder problems or ULD (Petty 2011). Since range of motion does not correlate conclusively with disability, this may not be sufficient (Olsen et al. 2000, Poitras et al. 2000, Kwak et al. 2005). The studies presented above suggest that appropriate management of patients with NS-NP requires thorough evaluation of ULD using suitable UL OM during the assessment and management process (McLean et al. 2011, Osborn and Jull 2013). This would enable physiotherapists to identify and quantify any ULD and include UL rehabilitation in the management plan, if indicated. Ongoing evaluation using the same UL OM would facilitate monitoring the progression of ULD and allow evaluation of the effectiveness of the ULD rehabilitation. Since the presence of a shoulder problem is known to increase the risk for recurrent, persistent or disabling problems in patients who have NS-NP (McLean et al. 2007), it is hypothesised that appropriate management of any ULD as part of a holistic management plan, in patients with NS-NP may help to improve the overall effectiveness of that management plan. However, there is no clear guidance regarding the availability and suitability of ULD OMs for patients with NS-NP.

Physiotherapists play a key role in the management of patients with NS-NP and this usually involves a multimodal approach to management, which incorporates a wide range of possible conservative treatment approaches. This could include active treatment approaches such as therapeutic exercise, the McKenzie method and patient education, and passive treatment approaches such as manual therapy, electrotherapy and acupuncture (Moffett and McLean 2006). Limited evidence suggests that UK-based physiotherapists rarely consider UL rehabilitation when managing their patients with NS-NP (McLean et al. 2010b, McLean et al. 2013). In addition, to date there is no

empirical evidence that has investigated current clinical UK physiotherapy practice in relation to the measurement or management of patients with NS-NP.

1.4 The Single Arm Military Press (SAMP) test

Although there are a variety of measures that evaluate UL functional capacity in patients with NP, the SAMP test, as far as the author is aware, is the only performance-based instrument that was designed to specifically measure ULD in female patients with NS-NP (McLean et al. 2010a). Female patients with NS-NP were the focus of the SAMP test because they are more commonly affected by NP and they tend to have higher rates of using physiotherapy services (Hogg-Jonson et al. 2008, Cote et al. 2008, Sahin et al. 2008, Hoy et al. 2010, Hoy et al. 2014).

The SAMP test performance consists of tasks of functional relevance (i.e. carrying, lifting and repetitive overhead activity), which challenge the UL (neck, shoulder, elbow, arm and hand) and often impaired in patients with NS-NP (McLean et al. 2011, Osborn and Jull 2013). It uses readily available and inexpensive equipment (one dumbbell) and it is very easy to score (repetition count within 30 seconds). The SAMP test is conducted with the patient in the standing position with their feet positioned at shoulder width. The patient is asked to carry a dumbbell and to lift it, using their dominant hand/other hand, to shoulder level (see Figure 1.1A). The patient is requested to raise their hand with the dumbbell directly overhead by extending through the elbow (see Figure 1.1B) and repeat this process as fast and as frequently as possible for 30 seconds (McLean et al 2010a). These tasks evaluate the strength and endurance of the UL, with expectation that the difficulty in sustaining overhead activity within 30 seconds would discriminate between NS-NP patients with varying degrees of ULD. Therefore, the SAMP test performance has a greater likelihood of accurately identifying and quantifying any UL functional limitation in patients with NS-NP (Curb et al. 2006, Pinheiro et al. 2016). In addition, the SAMP test is a simple test that can be efficiently administered by physiotherapists, clinicians, and/or individuals with varying experience in any setting using the minimum of equipment (i.e. a single dumbbell) in less than 2 minutes. Therefore, it has the capacity and characteristics to be very useful for use in day-to-day busy clinical practice as well as research practice. The SAMP test was developed and validated in a series of preliminary studies and demonstrated excellent reliability and validity. However, these studies were conducted on female, non-patient populations using a 3-kg hand weight in the SAMP's practical application, which is

considered to be unsuitable (too heavy) for the patient group. Therefore, the suitability, measurement and practical properties of the SAMP test in patient populations with NP are still unclear.

Fig 1A Fig 1B

Figure 1.1 SAMP Test Protocol

1.5 Summary and research questions

In summary, there is a strong association between NS-NP and ULD, and optimal management of NS-NP might incorporate UL rehabilitation and therefore should include evaluation of the UL functional capacity using a suitable ULD OM. However, there is no clear guidance regarding the availability and suitability of UL OMs for patients with NS-NP. In addition, there is no information available on how UK physiotherapists measure or rehabilitate their patients with NS-NP and ULD. The SAMP test is a potentially useful performance-based OM that was designed to measure ULD in patients with NS-NP. It is simple, quick, inexpensive, easy to administer in any setting and has the characteristics to be very useful in clinical practice as well as in research practice, however it still requires validation in patient groups with NS-NP. Therefore, this thesis was designed to answer the following research questions (RQ):

1. What are the measurements and practical properties of all available ULD OMs that have been developed or validated for patients with NS-NP?
2. What are UK physiotherapists' current measurement and management strategies for patients with NS-NP?

3. What are the acceptability and feasibility of the SAMP test in female patients with NS-NP?
4. What are the reliability and validity of the SAMP test in female patients with NS-NP and healthy subjects?

This work provides empirical evidence regarding the suitability of currently available upper limb OMs for patients with NS-NP, which reveals the need for a suitable upper limb OM for patients with NS-NP (RQ1) and supports the selection of the SAMP test for further adequate validation in patients with NS-NP (RQ 3 & 4). In addition, this work provides an insight into UK physiotherapists' use of measurement and management strategies for NS-NP and ULD (RQ 2) and ultimately provides a measure (SAMP test) which might facilitate that measurement and management.

1.6 Philosophical framework: Positivism

Research philosophy is a system of theories, ideas, principles and assumptions about the development of knowledge (Klee 1997). Shepard and colleagues (1993) observed that at every stage of research, a number of types of assumption will be made. These include assumptions about reality and its nature (ontological assumptions); human knowledge, what is considered acceptable knowledge and what kinds of contribution to knowledge can be made (epistemological assumptions); and the role of values and ethics within the research process (axiological assumptions). It is recognised that consistent and well-planned assumptions will constitute a robust research philosophy, which in turn will facilitate identifying the most appropriate methodological approaches that will answer the research question comprehensively (Crotty 1998). Philosophical approaches are scattered between two opposing extremes: positivism and phenomenology (Tashakkori, and Teddlie 2010). Positivistic researchers often aim to discover objective reality that can be answered by formulating and testing one or more testable hypotheses that reflect anticipated answers to questions about the relationship between cause and effect (Phillips 1987). The main assumptions underlying positivism are that the phenomenon need to be measured; verification or hypotheses testing requires deductive processes; and therefore, the key methodological approach is experimentation via direct manipulation and observation (Trochim 2002). Conversely, the phenomenologist is the researcher who tries to understand human activity from the perspective of the individual being studied (Cohen 1987). The main assumptions underlying phenomenology are that reality is socially constructed by individual and thus multiple realities exist; understanding the unknown phenomenon requires inductive processes; and therefore,

the key methodological approach is exploration of pure subjectivity using qualitative methodologies (Landsheere 1988).

Previous research on measuring ULD in patients with NS-NP has been conducted within a positivist paradigm underpinned by an objectivist epistemology. In addition, the objective research focussing on exploring the clinical measurement and management of ULD in patients with NS-NP fall within the positivist philosophical framework. Therefore, this thesis adopted positivistic ontological position based on the fact that ULD can be precisely and accurately measured using OMs in all populations with NS-NP, which demonstrates that there is a single and external reality. This in turn led to adopting an objectivist epistemological approach and incorporating deductive reasoning research. A positivist philosophical approach utilising quantitative methods and designs makes: 1) exploration of the suitability of all available ULD OMs for patients with NS-NP; 2) gaining an insight into UK physiotherapists' use of measurement and management strategies for NS-NP and ULD; and 3) further validation of the performance-based OM (SAMP) test precisely and accurately possible. Therefore, a positivism philosophical framework of inquiry is appropriate for this thesis. Consequently, the systematic review method, quantitative survey design and quantitative evaluation of the measurement and practical properties of the SAMP test were identified as the most appropriate to comprehensively answer the research question.

1.7 Research ethics and governance approval

Research ethics and governance permission from Sheffield Hallam University Research Ethics and Governance Committee to develop this programme of research was gained on 06/09/2013. The specific ethical concerns to each chapter of this thesis have been discussed within the individual chapters of the thesis. A single ethical approval was sought and gained for the survey in chapter four on 17/02/2015 from the Health & Wellbeing Faculty Ethics Committee at Sheffield Hallam University. The letter of ethics approval is at appendix 4. The validation studies in chapters five and six were conducted on Egyptian female patients with NS-NP and healthy subjects. Initially research approval was sought for these studies on 08/06/2015 and gained on 06/07/2015 from Tanta Universal Teaching Hospital following submission of an application form, research protocol, questionnaires, research information sheet, research participants consent form, CV for the researcher and CV for the director of study. This allowed

Tanta Universal Teaching Hospital (Rheumatology and Physical Therapy Department) to participate in the research. Subsequently, ethical approval was sought and gained on 26/10/2015 from the Health & Wellbeing Faculty Ethics Committee at Sheffield Hallam University to conduct these studies in Egypt. Letters of ethics approval are at appendix 7 and 8.

Overall, the main ethical concern in this thesis relates to conducting these validation studies in Egypt. Egyptian research ethics guideline exist; however, it has numerous deficiencies in their stated protections to research participants in relation to clinical and experimental research (Alahmad et al. 2012). Therefore, these studies were conducted in accordance with the UK guideline for health and social care research in order to meet international ethical standards. The following principles were considered when conducting the research: (1) the safety and well-being of participants; (2) competence of staff involved in conducting the research; (3) integrity, quality and transparency of the research; (4) research protocol; (5) benefits and risks for the individual participants; (6) gaining approval before commencing; (7) information about the research; (8) providing choice to participants without reprisal; (9) and respect of privacy as all information collected was recorded, handled and stored appropriately so that it can be utilised while the confidentiality of participants remain protected. In addition, religious and cultural issues were considered when assessing and testing Muslim female patients.

1.8 Structure of the thesis

The PhD project was designed in five parts using four distinct research methods (see Figure 1.2 below). In order to answer the research questions identified in section 1.5 above. Part one was a systematic review that aimed to identify, summarise and critically examine all available studies on the measurement and practical properties of all available ULD OMs that were used for patients with NP and make recommendations about those that are suitable for use in clinical practice and research (RQ 1). Part two was a literature review to explore current evidence-based management practice within the scope of physiotherapy for patients with NS-NP. The findings of this literature review and the systematic review were used to inform the development of the subsequent national survey of UK physiotherapists' measurement and management of patients with NS-NP (Part three) (RQ 2). In the context of this programme of research, the particular area of interest was related to the utilisation of OMs in the assessment and during the management of NS-NP as well as whether or not UK physiotherapists

consider ULD rehabilitation when managing their patients with NS-NP. Part four was a cross-sectional study that investigated the acceptability and feasibility of the SAMP test at lower weight (½-kg, 1-kg, 1½-kg) in Egyptian female patients with NS-NP, identified from the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital (Egypt) (RQ 3). Part five was a validation study that investigated the reliability and validity of the SAMP test in Egyptian female patients with NS-NP and healthy subjects (RQ 4). The structure and content of the reminder of thesis are summarised below.

Chapter 2: Measures of upper limb function for people with neck pain: A systematic review of the measurement and practical Properties.

This chapter identifies and reviews the measurement and practical properties of all available ULD OMs that were developed or validated for patients with NP. This chapter addresses RQ 1.

Chapter 3: Evidence of the currently recommended treatment approaches for nonspecific neck pain: A literature review.

This chapter describes a literature review exploring current evidence-based management practice within the scope of physiotherapy for patients with NS-NP. The findings of this literature review were used to inform the development of the subsequent survey.

Chapter 4: Physiotherapy management of patients with non-specific neck pain: A national survey of current UK practice.

This chapter describes a national survey investigating UK physiotherapists' measurement and management of patients with NS-NP to establish the current utilisation patterns of OMs in the assessment and during the management of NS-NP as well as the treatment approaches that are most often used alongside ULD rehabilitation. This chapter addresses RQ 2.

Chapter 5: Measuring upper limb disability in neck pain population: Evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test.

This chapter describes a study investigating the acceptability and feasibility of the SAMP test from both the patients' and clinicians' perspectives using lower weight (½-kg, 1-kg, 1½-kg) in the practical application of the SAMP test. This chapter addresses RQ 3.

Chapter 6: Measuring upper limb disability in Egyptian female patients with nonspecific neck pain: Evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test.

This chapter describes a study investigating the reliability (inter- and intra-rater and measurement error) and the construct validity (convergent and discriminate) of the SAMP test in patient populations with NS-NP. This chapter addresses RQ 4.

Chapter 7: Summary, discussion and conclusion

This chapter summarises the key findings of the thesis and provides recommendations about how to prevent ULD in patient populations with NS-NP.

Figure 1.2 Structure of the PhD Project

1.9 Publications and presentations

1.9.1 Publications in peer reviewed journals

The following peer-reviewed papers, incorporating research from this PhD project have been published:

- ALRENI. A., HARROP D., GUMBER A., MCLEAN S. (2015). Measures of upper limb function for people with neck pain: a systematic review of the measurement and practical properties (protocol). Systematic Review, 4 (43), 0034-39.

- ALRENI A., HARROP D., LOWE A., POTIA T., KILNER K., MCLEAN S. (2017). Measure of upper limb function for people with neck pain. A systematic review of the measurement and practical properties. Musculoskeletal Science and Practice, 29, 155-163.

1.9.2 Conference presentations

- ALRENI, A. (2016). Measures of upper limb function for people with neck pain: A systematic review of measurement and practical properties. Poster presentation at IFOMPT, Glasgow, Jul 2016.

- ALRENI, A. (2017). Managing non-specific neck pain: A national survey of current UK physiotherapy practice. Oral presentation at The Annual General Meeting of The Society for Back Pain Research, Northampton UK, Nov 2017.

- ALRENI, A. (2017). Outcome measures utilisation in managing non-specific neck pain: A national survey of current physiotherapy practice in the UK. Poster presentation at Physiotherapy UK (CSP Conference & Trade Exhibition), Birmingham UK, Nov 2017.

- ALRENI, A. (2018). Measuring upper limb disability in patients with neck pain: Evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test. Poster presentation at Physiotherapy UK (CSP Conference & Trade Exhibition), Birmingham UK, 2018.

- ALRENI, A. (2018). Measuring upper limb disability in patients with neck pain: Evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test. Oral presentation at Physiotherapy UK (CSP Conference & Trade Exhibition), Birmingham UK, 2018.

1.9.3 Manuscript under preparation

- ALRENI, A, HARROP, D., KILNER, K., DEMACK, S., MCLEAN, S. (2018). Outcome measures utilisation in managing non-specific neck pain: A national survey of current physiotherapy practice (Manuscript under preparation to be submitted to an appropriate Journal)

- ALRENI, A, HARROP, D., KILNER, K., DEMACK, S., MCLEAN, S. (2018). Managing non-specific neck pain: A national survey of current UK physiotherapy practice (Manuscript under preparation to be submitted to an appropriate Journal)

- ALRENI, A, HARROP, D., KILNER, K., MCLEAN, S. (2019). Managing nonspecific neck pain: Reporting the Social Media Strategy that was used in the UK national survey of neck pain recruitment and administration (Manuscript under preparation to be submitted to an appropriate Journal)

- ALRENI, A, HARROP, D., KILNER, K., MCLEAN, S. (2019). Measuring upper limb disability in neck pain population: Evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test (Manuscript under preparation to be submitted to an appropriate Journal)

- ALRENI, A, HARROP, D., KILNER, K., MCLEAN, S. (2019). Measuring upper limb disability in neck pain population: Evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test (Manuscript under preparation to be submitted to an appropriate Journal).

The chapter as follows addresses the first research question and describes the systematic review completed to identify, summarise and critically examine all available studies on the measurement and practical properties of all available OMs that have been developed or validated to measure ULD in patient populations with NS-NP.

Chapter 2: Measures of upper limb function for people with neck pain: A systematic review of the measurement and practical properties.

2.1 Introduction

This chapter forms part one of the thesis and describes a narrative systematic review completed to answer research question one of the PhD project (see section 1.5): to identify, summarise and critically examine all available studies on the measurement and practical properties of all available OMs developed or validated to measure ULD for patients with NP. The aim and objectives of the systematic review are summarised alongside definitions of the measurement properties of an OM according to the "**CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments" (COSMIN) checklist (Mokkink et al. 2010a, Terwee et al. 2012). The methods used in the systematic review are described and then the results of the review are presented and discussed.

2.1.1 Aim

The aim of this systematic review was to explore all existing outcome measures that were developed or validated to measure ULD in patents with NP. The Objectives were to:

1. Identify all OMs used to measure ULD in patients with NP
2. Summarise and critically appraise the methodological quality of all available studies on the measurement and practical properties of the identified OMs.
3. Provide recommendations about the relevance and suitability of OM for application in clinical practice and research.

2.1.2 Measurement properties of outcome measure

Clinically, OMs are used for a variety of purposes: (1) before interventions for screening of symptoms/function, capturing the aspects of health that matter most to patients, classifying patients into meaningful sub-groups, assisting clinical reasoning and setting treatment goals (diagnosis and prognosis) (Lansky et al. 1992, Kramer and Holthaus, 2006, Kyte et al. 2015), (2) during interventions to monitor condition progression and detect changes in pain/disability (Richard et al., 1992; Garland et al., 2003; Bot et al., 2007; Nordin et al., 2008) and (3) after interventions to determine the effectiveness, efficiency and cost-effectiveness of the interventions used and monitoring patient valued outcomes (CSP 2012, van Dulmen et al. 2017). The measurement properties (e.g. reliability, validity and responsiveness) of an OM should be adequate as any failures of these measurement properties would lead to imprecise evaluation and incongruous

decisions regarding the management. The measurement properties definitions from the COSMIN taxonomy study (Mokkink et al. 2010b) were used in this thesis as the foundation for providing definitions for terminology of the measurement properties.

Measurement properties of an OM are divided according to the COSMIN taxonomy study into three domains: (1) reliability, (2) validity and (3) responsiveness. Interpretability was considered to be sufficiently important by the COSMIN panel to be included in the COSMIN taxonomy despite that it is not a measurement property for quality testing (see Figure 2.1) (Mokkink et al. 2010b).

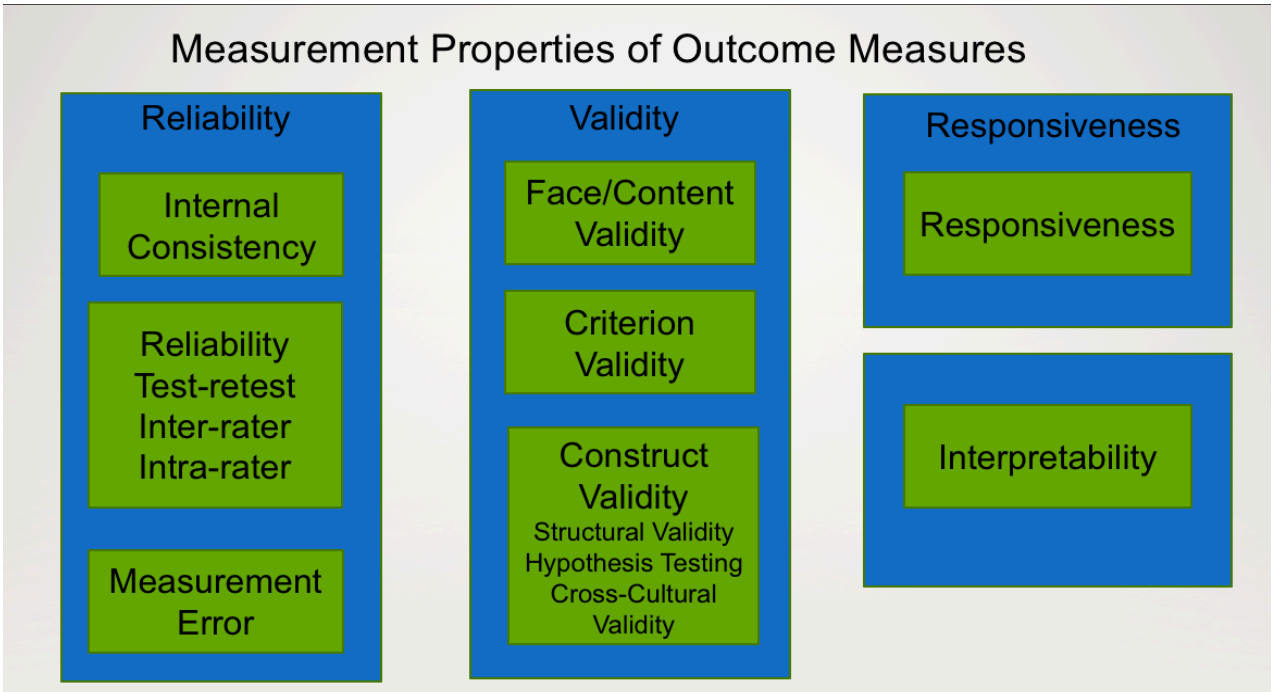


Figure 2.1 COSMIN Taxonomy (Mokkink et al. 2010b)

2.1.2.1 Reliability

It is a fundamental requirement that all OMs incorporated into clinical practice and research are reliable, and this is to ensure the accuracy of scores under different conditions when a patient is stable (de Vet et al. 2011). Reliability as a domain is defined by the COSMIN panel as "the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions (e.g. using different sets of items from the same multi-item measurement instrument (internal consistency), over time (test-retest), by different persons in the same occasion (inter-rater) or by the same person in different occasions (intra-rater)" (Mokkink et al. 2010b). Internal consistency, reliability and measurement error are the measurement properties associated with the reliability domain.

Internal consistency is defined as the extent to which items in a questionnaire are interrelated (Mokkink et al. 2010b). It is a measure of the extent to which items assess the same construct in a unidimensional scale of a multi-item instrument (de Vet et al. 2011). Cronbach's alpha is the parameter frequently used to assess the level of internal consistency, in which Cronbach's alpha values between 0.70 and 0.90 represent a well-accepted guideline of internal consistency (de Vet et al. 2011). Internal consistency is a redundant measurement property for objective or performance-based OMs.

Reliability as a measurement property is described as the proportion of the total variance in the measurement resulting in the consistency of the scores (Mokkink et al. 2010b). Reliability is considered the consistency of the results obtained from (test-retest, inter- and intra-rater) and expressed in correlations using the Interclass Correlation Coefficient (ICC) or Kappa (de Vet et al. 2011). The ICC range is between 0.0 to 1.0, where values close to 0.0 indicate poor reliability and ICC values close to 1.0 suggest high reliability (Portney and Watkins 2009).

Measurement error is defined as the error which is not attributed to true changes in the construct measured but resulted in the systematic and random error of a patient's score (Mokkink et al. 2010b). It is the absolute measurement error over repeated administration of the test when the patients are stable and it is represented by the Standard Error of Measurement (SEM), in which a low level of SEM indicates high levels of score accuracy and a high level of SEM indicates low levels of score accuracy (Vincent and Weir 2012).

2.1.2.2 Validity

Validity is an essential measurement property that should be possessed by an OM since it determines the true association between the OM and the construct of interest (de Vet et al. 2011). The domain validity is defined as the extent to which OMs truly measure the construct which they are expected to measure (Mokkink et al. 2010b). The validity domain, according to the COSMIN taxonomy, is divided into the three measurement properties as follows.

Content validity is defined as the extent to which the content items/tasks of an OM is an adequate reflection of the construct to be measured and examines the extent to which the constructs of interest are comprehensively represented by those items/tasks

(Mokkink et al. 2010b). Face validity which is considered an aspect of content validity in the respect that it concerns the degree to which an OM appears as though it is an adequate reflection of the construct being assessed. Content validity is assessed to ensure that the OM adequately represents the construct under study; this emphasises the importance of a good description of the construct to be measured and implies that OM items/tasks should be both relevant and comprehensive (de Vet et al. 2011). Relevance is assessed by answering the following three questions. First, do all items/tasks refer to the relevant aspect of the construct of interest? Second, are all items/tasks relevant to the study population (e.g. age, gender, disease characteristics, languages, countries, setting)? Third, are all items/tasks relevant for the purpose (e.g. discrimination "distinguish between patients in one occasion", evaluation "assess change over time" or prediction "predict future change") of the application of the OM? (Terwee et al. 2007).

Criterion validity is defined as the degree to which the scores of an OM are an adequate reflection of a gold standard (Mokkink et al. 2010b). Gold standard OMs which represent the true state of the construct of interest seldom exist in practice (de Vet et al. 2011). Patient-Reported Outcome Measures (PROMs) which always acquire subjective information, often lack a gold standard. However, in circumstance such as when one wants to develop a shorter questionnaire for a construct, when a long version of this questionnaire already exists (i.e. DASH and QuickDASH), the long version is considered an adequate gold standard for the shorter version (Mokkink et al. 2010b). For objective or performance-based OMs, a gold standard usually is an OM or reliable assessment criteria for the construct under study that has been accepted and is regarded by experts in the field as ideal to identify a condition and/or measure its severity (de Vet et al. 2011). Criterion validity is sub-divided by the COSMIN taxonomy to: (1) concurrent validity (that is the assessment by comparing the score of the OM under study and the gold standard at the same time, which is usually assessed for OMs to be used for evaluative and diagnostic purposes), and (2) predictive validity (that is, the assessment of whether the OM under study predicts the gold standard in the future, which is usually measured for OMs to be used in predictive applications) (Mokkink et al. 2010b). Assessing criterion validity requires comparing the scores of the OM under study with the scores obtained from the criterion OM. This is often determined by the level of measurement for the OM under study and the criterion. For example, Sensitivity and Specificity are adequate parameters when both OMs have a dichotomous outcome and are expressed by the same unit of measurement. The Receiver Operator Curve

(ROCs) is adequate when the OM under study has an ordinal or continuous scale with dichotomous criterion and different unit of measurement. Bland and Altman limits of agreement (LoA) or Interclass Correlation Coefficient (ICC) should be used when both OMs have a continuous scale and are expressed by the same units of measurement (de Vet et al. 2011).

Construct validity should be used to provide evidence of the validity of an OM when a gold standard of the construct to be measured is not available. It concerns the degree to which the scores of the instrument under study are consistent with clearly and a priori formulated hypotheses regarding the relationship with the scores of other instruments that should be measuring the same construct and it is often assessed using the Pearson correlation coefficient (Mokkink et al. 2010b). Construct validity is sub-divided by the COSMIN taxonomy into the following three properties. First, *structural validity*, which is defined as the extent to which the scores of an OM are an adequate reflection of the dimensionality of the construct being measured (Mokkink et al. 2010b). It is assessed using factor analysis to confirm the number of subscales presented in a questionnaire (de Vet et al. 2011). Consequently, it is relevant to PROMs and redundant for objective or performance-based OMs. Second, *hypothesis testing*, which is the basic principle of construct validity since it is described based on the idea that hypotheses are formulated about the relationship of scores on the OM under study and scores on other OMs measuring similar or dissimilar constructs, or differences on the OM scores between sub-groups of patients (Mokkink et al. 2010b). Hypotheses should be as specific as possible, formulated prior to data collection and reported together with their justification to allow for assessment of their plausibility (de Vet et al. 2011). Formulated hypotheses have then to be tested and assessed based on their expected level and direction of correlation with the comparative OMs (Terwee et al. 2007). Third, cross-cultural validity is the extent to which items in a questionnaire can mirror the performance of the same items when the questionnaire translated into another language or adapted to reflect the lifestyle of a different culture (Mokkink et al. 2010b). It is often assessed after the translation of a PROM questionnaire by evaluating the construct validity of the translated version. This is to examine whether the translated OM demonstrates the expected correlations with related constructs, and it has the capability to discriminate between the relevant sub-groups of patients. Consequently, this measurement property is relevant only to the PROMs and it is redundant for objective or performance-based OMs.

2.1.2.3 Responsiveness

The domain responsiveness is defined by the COSMIN taxonomy as the ability of an OM to detect changes over time in the construct being measured (Mokkink et al. 2010b). Consequently, evaluative OMs used in clinical practice and research practice should have the capability to detect and quantify changes in health status overtime in the construct of interest (de Vet et al. 2011). Responsiveness is assessed using the same methodological principle as validity since it is an aspect of validity, however the only difference is that responsiveness emphasises the validity of change scores, while validity emphasises the validity of single scores (Mokkink et al. 2010b). This implies that a longitudinal study is required, in which two measurements should be taken in order to calculate change scores and changes in the construct of interest are expected (i.e. at least some proportion of patients would improve or deteriorated). This is because if no change on the OM was detected, it would be difficult to determine whether this was because the patients really did not change, or if the OM was not responsive (de Vet et al. 2011).

2.1.2.4 Interpretability

Interpretability is defined by the COSMIN taxonomy as the degree to which clinician can assign qualitative meaning to an OM's quantitative scores (Mokkink et al. 2010b). It is not a measurement property, as reliability and validity, since it does not refer to the quality of an OM but it refers to what the scores on an OM means. Adequate interpretability of a score is necessary before considering the use of an OM in clinical practice and research (de Vet et al. 2011). Interpretability can be assessed by examining the distribution of scores, the occurrence of floor and ceiling effects and the availability of scores and change scores for the relevant sub-groups as well as the calculation of the Minimal Important Change (MIC) or the Minimal Important Difference (de Vet et al. 2011).

2.1.3 Practical properties of outcome measure

Practical properties are those related to the practicality and burden of patients (patients' acceptability and feasibility), and the practicality and burden of clinicians (clinicians' acceptability and feasibility) as well as precision of an OM (Selby and Robertson 1987, Erikson et al. 1995, Kessler and Mroczek 1995). These practical properties were first

highlighted in 1998 as essential properties which should be possessed by an OM that is considered for use in clinical practice and research (Fitzpatrick et al. 1998).

Patients' acceptability is defined as the ability and willingness of a patient from the target population to complete questions or tasks related to an OM (Fitzpatrick et al. 1998). Meanwhile, patients' feasibility (burden) is described as the time and effort required from a patient to complete questions/tasks and the proportion of patients who find these questions/tasks difficult or impossible to complete for any reason (Selby and Robertson 1987). Patients' acceptability and feasibility comprised two main components. First, reasons for non-completion: if a patient was unable to complete the questions in a PROM or tasks in a performance-based OM because of difficulties or distress, this is an indication of unacceptability and/or unfeasibility of this OM unless other reasons such as health status deterioration or other disabilities were involved (Fitzpatrick et al. 1998). Second, completion time is often considered to be a determinant of an instrument's acceptability and feasibility (the shorter the time it takes to complete, the more acceptable and feasible the OM is to the patient) (Nelson et al. 1990).

Clinicians' acceptability is frequently related to the difficulty clinicians encounter during the administration of an OM, such as the length and complexity of the overall testing procedure. Clinicians' feasibility (burden) is related to the resources required from the clinicians to compete the testing procedure, and this includes the time and cost of administration, speed and ease of scoring and feedback of information and interpretation (Fitzpatrick et al. 1998). This suggests that brevity, simplicity in administration and ease of the scoring system alongside free access to OMs is an indication of greater clinicians' feasibility (Read et al. 1987, Feeny and Torrance 1989, Nelson et al. 1990). Both patients' and clinicians' acceptability and feasibility of an OM are essential properties and should be established prior to the testing of other measurement properties such as reliability, validity and responsiveness (Sprangers et al. 1993). They can be assessed by obtaining the qualitative opinion of patients as well as clinicians regarding their experience with the OM under study directly after administration (Fitzpatrick et al. 1998).

2.2 Methods

2.2.1 Design

narrative two-phase systematic review was undertaken to explore all OMs that were developed or validated to assess ULD in patient populations with NP. Phase one identified all OMs that have been used to assess ULD in patients with NP. Phase two identified all available studies investigating the measurement and practical properties of the identified OMs. The methodological quality of the developmental and/or evaluative studies of those identified OMs were assessed against the "COnsensus-based Standards for the selection of health Measurement INstruments" (COSMIN) checklist (Mokkink et al. 2010a, Terwee et al. 2012). The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guideline for systematic reviews was followed in reporting this study (Liberati et al. 2009, Moher et al. 2009).

2.2.2 Phase one – identification of measures

2.2.2.1 Data source and search strategy

The bibliographic databases as follows were searched from their inception until March 2016: Allied and Complementary Medicine Database (AMED) (OvidSP), CINAHL Complete (EBSCO), Cochrane Library (Wiley), MEDLINE (EBSCO), PubMed (US National Library of Medicine), PsycINFO (ProQuest), SPORTDiscus (EBSCO), Web of Science (Thomson Reuters).

The search strategy in this phase of the study comprised terms relating to ULD and NP. These terms were combined with Boolean logic terms. Other terms were incorporated to limit the search to OMs, psychometric properties or measurement properties. The searches were undertaken in February and March 2016. All search terms were looked for in the title and abstract fields and controlled vocabulary terms were used where available. The Boolean operators AND and OR were used, alongside truncation, phrase searching and proximity operators. The search strategy for MEDLINE (EBSCO) is as follows (see Box 2.1). The search syntax detailed below were adapted for use on other information resources used in the search.

Box 2.1: Search strategy for phase 1

("upper limb"[ti,ab] OR "upper extremity"[ti,ab] OR function* [ti,ab] OR dysfunction*[ti,ab] OR abilit*[ti,ab] OR disabilit*[ti,ab] OR capacity*[ti,ab] OR disorder*[ti,ab] OR problem*[ti,ab] OR pain*[ti,ab] OR deficit*[ti,ab] AND neck[ti,ab] OR "cervical spine"[ti,ab] OR cervicogenic*[ti,ab] OR pain*[ti,ab] OR function*[ti,ab] OR dysfunction*[ti,ab] OR abilit*[ti,ab] OR disabilit*[ti,ab] OR problem*[ti,ab] OR disc*[ti,ab] OR "degenerative disc"[ti,ab] OR degeneration*[ti,ab] OR disease*[ti,ab] OR disorder*[ti,ab] OR deficit*[ti,ab] AND "outcome measure*" n5[ti,ab] OR "outcome assessment*"[ti,ab] OR psychometr*[ti,ab] OR clinimetr*[ti,ab] OR "observer variation*"[ti,ab] OR reproducib*[ti,ab] OR reliab*[ti,ab] OR unreliab*[ti,ab] OR valid*[ti,ab] OR discriminant*[ti,ab] OR coefficient*[ti,ab] OR correlation*[ti,ab] OR selection*[ti,ab] OR reduction*[ti,ab] OR agreement*[ti,ab] OR precision*[ti,ab] OR imprecision*[ti,ab] OR test-retest*[ti,ab] OR interrater*[ti,ab] OR intrarater*[ti,ab] OR inter-rater*[ti,ab] OR intra-rater*[ti,ab] OR kappa*[ti,ab] OR "minimal important change*"[ti,ab] OR "multitrait scaling analysis*"[ti,ab] OR "factor analysis*"[ti,ab] OR "known group*"[ti,ab] OR responsive*[ti,ab].

Note: (ti) = title field, (ab) = abstract field, (/) = MeSH, asterisk (*) denotes any character, ("") = phrase search, (n5) = adjacency within five words.

2.2.2.2 Study selection

Inclusion criteria

All studies yielded from the literature search were eligible for inclusion in this review without restriction of study design or publication date provided the article: (1) was a full-text original primary quantitative study (e.g. clinical trials, observational studies, case-controlled studies or case studies), (2) was published in the English language, (3) involved adult ≥ 18years of age with NP, which is defined here as "pain perceived as arising from anywhere within the region bounded superiorly by the superior nuchal line, inferiorly by the transverse line through the tip of the first thoracic spinous process, and laterally by the sagittal planes tangential to the lateral border of the neck (Merskey and Bogduk, 1994), and (4) contained at least one OM to measure ULD. For the purpose of this study ULD is defined as the difficulties or limitation an individual may have when executing tasks/activities using their ULs such as carrying, lifting and overhead activities (ICF 2001).

Exclusion criteria

Articles were excluded if: (1) they did not use primary quantitative data, (e.g. systematic reviews, meta-analysis, qualitative studies, reportage or opinion pieces), (2) they did not include at least one OM to measure ULD, or (3) involved patients with disorders other than NP.

Screening

After completion of the search process, the results were initially reviewed by one reviewer (ASEA) to exclude any duplication and obviously irrelevant studies. This was followed by a two-phase screening strategy to identify the studies to be reviewed. Firstly, two reviewers (ASEA and AL) independently screened the title and abstract of the articles retrieved against the inclusion and exclusion criteria and selected all potentially relevant studies. Finally, the full-text articles were retrieved and the aforementioned reviewers independently screened each of the retrieved articles to further determine their eligibility for inclusion in this review. In the case of a disagreement between the two reviewers as to whether an article should be included or excluded, a consensus was sought through discussion, and if required a third reviewer (SMc) made the final decision. Reference lists of all included studies were also scrutinised independently by the two reviewers to identify additional relevant articles.

2.2.3 Phase two – identification of the developmental and/or evaluative studies

A second search was performed, using the databases identical to those searched in phase one. The name of each OM identified in phase one was searched for using all fields search function and was used to identify all articles related to the development or validation of the measurement and practical properties of this OM. A sensitive search filter (Terwee et al. 2009), was used to locate articles reporting the measurement and practical properties of each identified OM. Furthermore, the authors and/or developers of specific OMs were contacted to request additional published and/or unpublished evidence of measurement evaluation.

2.2.4 Data extraction

A data extraction form (Appendix 1) informed by earlier reviews from Haywood et al. (2013), Haywood et al. (2014) and the COSMIN checklist (Mokkink et al. 2010a, Terwee et al. 2012) was used to capture study specific (population, intervention, and setting) and measurement specific information: reliability (internal consistency, test-retest, intra-/inter tester, measurement error), validity (face/content, structural validity (dimensionality), construct validity (evidence of explicit hypothesis testing, discriminant/discriminative), criterion validity (concurrent, predictive), responsiveness (criterion approach, construct approach), interpretability (for example, evidence of minimal important change), data precision (data quality, end effect), and evidence of where Item Response Theory (IRT) models where applied. Extraction of practical properties included acceptability (relevance and respondent burden) and feasibility

(clinician burden, including cost, time to complete/score). The extent of patient involvement in measurement development and/or application was also sought (Haywood et al. 2014). Two reviewers (ASEA and TP), independently performed the data extraction for all included studies. In the case of disagreement about a study, a consensus was reached between the two reviewers via discussion. A third reviewer (SMc) was available to make the final decision, if necessary.

2.2.5 Quality assessment of studies

Articles that were included for methodological quality, data analysis and data synthesis were those related to the development and/or validation of the measurement and practical properties of all available ULD OMs for patient populations with NP. Two reviewers (ASEA and TP), independently performed the methodological quality of studies selected for inclusion in this review. In case of disagreement, a consensus was reached through discussion. A third reviewer (SMc) was available to make the final decision, if required. Each identified OM was evaluated for its development or validation methodology, and measurement and practical properties. The methodological quality assessment was undertaken using the COSMIN checklist (Mokkink et al. 2010a, Terwee et al. 2012).

2.2.5.1 Rational for COSMIN

The COSMIN checklist (Mokkink et al. 2010a, Terwee et al. 2012) was used to critically appraise the quality of all studies included in this review. The checklist is a comprehensive and rigorous quality assessment tool, developed specifically to focus on the measurement properties and methodological quality of health-related OMs. It also incorporates a standardised rating system alongside multilevel grading for each measurement property. In addition, provides an overall quality rating for the methodological quality of a study in relation to each measurement property being assessed within that study. The COSMIN associative taxonomy study (Mokkink et al. 2010b) facilitates agreement between reviewers when a measurement property is reported using different terminology across multiple studies. It provides extensive guidelines to facilitate interpretation of items and score levels for each measurement property box. Further, it provides detailed standards regarding adequacy of design and statistical methods within studies evaluating the measurement properties of health-related OMs. Finally, the COSMIN checklist provides a grading for each measurement property rather than the OM as a whole, which informs decision-making regarding specific limitations of an OM.

2.2.5.2 COSMIN checklist

The COSMIN checklist (Mokkink et al. 2010a, Terwee et al. 2012) is a four-point scale (excellent, good, fair or poor) comprising twelve boxes out of which nine boxes are used to assess the methodological quality of the measurement properties of an OM (Terwee et al. 2012). In each measurement property box, there are a variety of criteria that are rated on the aforementioned 4-point rating scale. These include criteria such as sample size, the methods used to manage missing data and the statistical analysis used. The internal consistency property comprises 11-items, reliability 14-items, measurement error 11-items, face/content validity 5-items, structural validity 7-items, cross-cultural validity 15-items, hypothesis testing 10-items, criterion validity 7-items, and responsiveness comprises 18-items. A study's methodological quality is rated for each measurement property evaluated within the study and determined by the lowest rate "worst score counts". For example, the methodological quality of a measurement property is considered *excellent* if all criteria related to that property are adequate and rated excellent. However, the methodological quality of a measurement property will be rated poor if at least one criterion related to that property was inadequate and rated as poor (Appendix 2) (Terwee et al 2012).

2.2.6 Data analysis

Data was qualitatively synthesised using a best evidence synthesis to determine the overall quality and acceptability of each identified measure (Haywood et al. 2013, Haywood et al. 2014). Different studies on the measurement properties of each identified measure were summarised by combining their results on: (1) the number of studies in which the measurement property was assessed, (2) their methodological quality (COSMIN score), and (3) the consistency of the result for each study was also examined and considered positive (+), negative (-) or indeterminate (?) following the criteria reported by (Terwee et al. 2007) (see Table 2.1). This was presented alongside the level of evidence suggested by the Cochrane Back Review Group, in which the possible level of evidence for a measurement property is "strong", "moderate", "limited", "conflicting" or "unknown" (van Tulder et al. 2003, Furlan et al. 2009) (see Table 2.2). This level of evidence strategy has been employed by multiple systematic reviews of OMs and is now established practice (Schellingerhout et al. 2011, Schellingerhout et al. 2012). The methodological quality ratings for the measurement properties of each identified instrument were accompanied by the strength of the results

and the level of evidence in order to enable inference on the relative robustness of evidence for each available instrument.

Table 2.1: Quality criteria for measurement properties (Terwee et al. 2007).

Property	Rating†	Quality Criteria
Reliability		
Internal consistency	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
	?	Dimensionality not known OR Cronbach's alpha not determined
	-	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Reliability	+	ICC / weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80
	?	Neither ICC / weighted Kappa, nor Pearson's r determined
	-	ICC / weighted Kappa < 0.70 OR Pearson's r < 0.80
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined
	-	MIC ≤ SDC OR MIC equals or inside LOA
Validity		
Content validity	+	All items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement AND the questionnaire is considered to be comprehensive
	?	Not enough information available OR no target population involvement
	-	Not all items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement OR the questionnaire is considered not to be comprehensive
Construct validity		
-Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	-	Factors explain < 50% of the variance
- Hypothesis testing	+	Correlations with instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlations with related constructs are higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlations with instruments measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlations with related constructs are lower than with unrelated constructs
- Cross-cultural validity	+	No differences in factor structure OR no important DIF between language versions
	?	Multiple group factor analysis not applied AND DIF not assessed
	-	Differences in factor structure OR important DIF between language versions
Criterion validity	+	Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥ 0.70
	?	No convincing arguments that gold standard is "gold"
	-	Correlation with gold standard < 0.70
Responsiveness		
Responsiveness	+	Correlation with changes on instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70 AND correlations with changes in related constructs are higher than with unrelated constructs
	?	Solely correlations determined with unrelated constructs
	-	Correlations with changes on instruments measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlations with changes in related constructs are lower than with unrelated constructs

Notes: MIC: Minimal Important Change, SDC: Smallest Detectable Change, LOA: Limits of Agreement, ICC: Interclass Correlation Coefficient, AUC: Area Under the Curve, (+) = positive rating, (-) = negative rating, (?) = indeterminate rating.

Table 2.2: Level of evidence for the overall quality of measurement property (van Tulder et al. 2003, Furlan et al. 2009).

Level	Rating†	Criteria
strong	**+++ or --**	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
moderate	**++ or --**	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
limited	**+ or -**	One study of fair methodological quality
conflicting	**+/-**	Conflicting findings
unknown	**?**	Only studies of poor methodological quality

(+) = positive result, (-) = negative result, (?) = indeterminate results.

2.3 Results

2.3.1 Phase one

The search strategy in this phase resulted in a total of 1382 unique records being identified from the database searches, reducing to 982 after the removal of duplicates. Following the title and abstract screening process another 928 articles were excluded. The full-text of 54 articles were retrieved and reviewed against the inclusion and exclusion criteria. This resulted in the exclusion of 5 articles (not primary quantitative), 1 article (foreign language study), 21 articles (other than NP population), 8 articles (the OM does not measure ULD) and 14 articles (not OM's developmental or evaluative study). Screening the reference lists from the five retained articles resulted in 15 additional potentially relevant articles, of which one article met the inclusion criteria for this review. From this phase, six developmental and/or evaluative articles for five clearly described and reproducible ULD OMs for NP patients were included in the review.

2.3.2 Phase two

Evidence for the measurement and practical properties were sought for those identified five OMs in phase one. However, the database searches did not uncover any new records. Contacting the developers of specific measures resulted in six additional articles, of which one was excluded (not OM's developmental or evaluative study) and five of these were retained for inclusion in the review.

2.3.3 Results from phase one and phase two

In total, 11 articles on the developmental/evaluative of the five instruments were included in this review. Figure 2.2 shows the phase one and phase two outcomes at each

stage of selection and screening process as well as the reasons for exclusions. Since there was 97% agreement between the two reviewers (ASEA and AL) regarding the inclusion and exclusion of studies and consensus was reached through discussion for the remaining 3%, the third reviewer (SMc) was not used. These 11 articles provide evidence for five clearly defined and reproducible outcome measures of upper limb disability in the context of neck pain. Three are patient-reported questionnaires: The Disability of Arm, Shoulder and Hand (DASH) (Hudak et al. 1996); the Quick Disability of Arm, Shoulder and Hand (QuickDASH) (Beaton et al. 2005); and the Neck and Upper Limb Index (NULI) (Stock et al. 2003). One is clinician-reported: The Shoulder Functional Assessment (SFA) (Lomond and Cote 2009). One is a performance-based test: The Single Arm Military Press (SAMP) (McLean et al. 2010).

The general characteristics of the 11articles are presented in Table 2.3. A summary of the quality of the measurement properties that were tested in each study is presented in (Appendix 3). The methodological quality of each study per measurement property is presented in Table 2.4. A synthesis of the results for each instrument, alongside their level of evidence is presented in Table 2.5. Since there was 95% agreement between the two reviewers (ASEA and TP) on the individual COSMIN items reviewed and consensus was reached through discussion for the remaining 5%, the third reviewer (SMc) was not used. A summary of the measurement properties for each identified instrument follows.

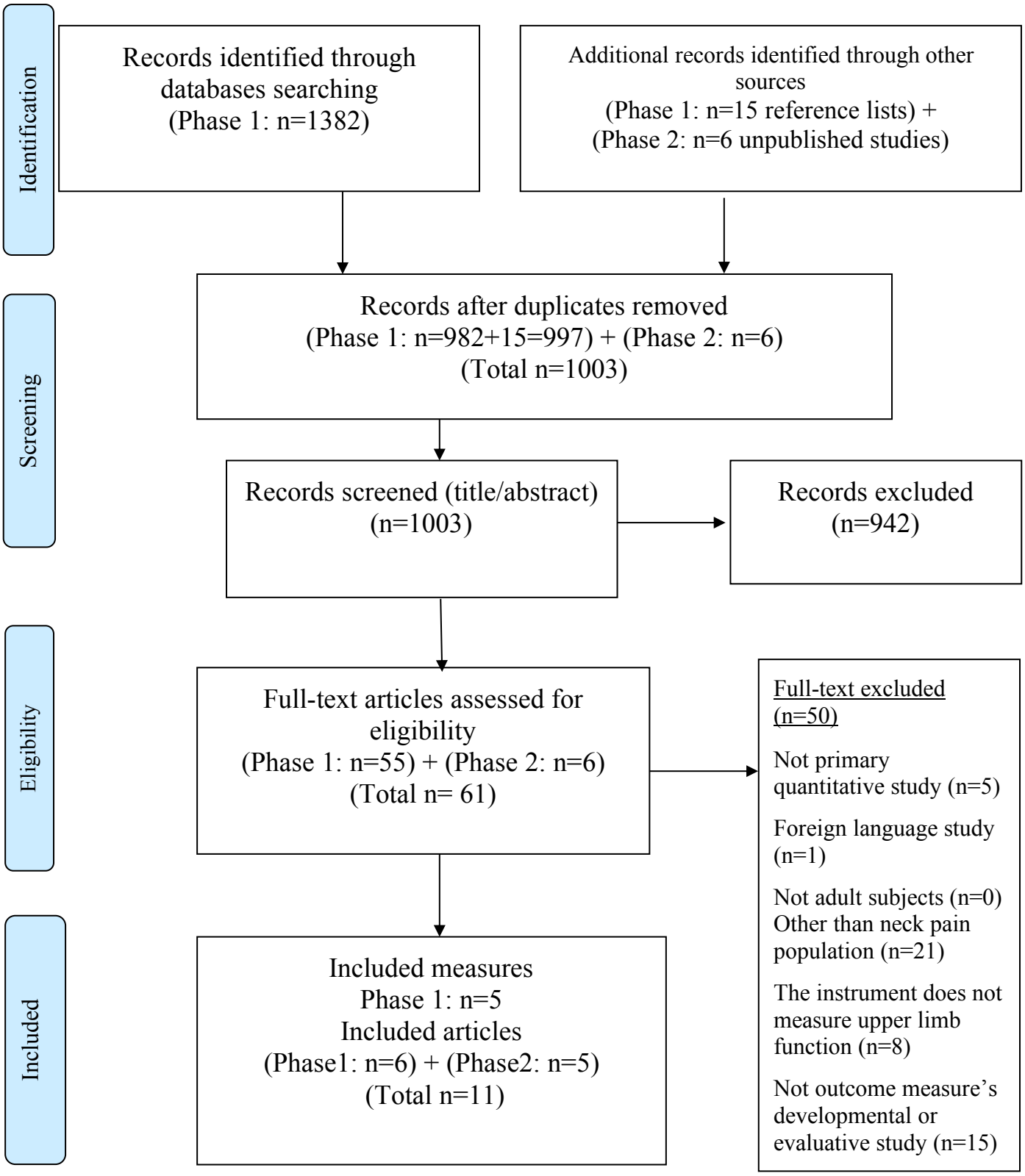


Figure 2.2 PRISMA Flow Chart of phase One and Phase Two.

Table 2.3: Characteristics of the included studies

Study	Sample Size	Mean age ± SD (range)	Population	Country	Setting	Recruitment methods	Outcome measures used in the study	Measurement property assessed
Huisstede (2009)	N=679	44.4 ± 11.4 (18-64)	Neck, shoulder, and/or arm pain	Netherland	Dutch General Practices (GPs)	Convenience	DASH SF-12 Severity of complaint Persistence of complaint	Hypothesis testing Responsiveness
Mehta (2010)	N=66	40.6 ± 14.2	Neck pain with/without arm pain, headache and whiplash disorders level 2&3	Canada	Canadian Physical Therapy Clinics	Convenience	DASH QDASH NDI VAS CSOQ	Hypothesis testing Concurrent validity
Fan (2008)	N=733	Total Sample: N=733 39.5 ± 0.05 Clinical-Cases: N=231 43.2 ± 0.7 Symptomatic Only: N=175 39.3 ± 0.8	Neck Or Upper Extremity Musculoskeletal Disorders (UEMSDs)	USA	Workplace walkthrough at 12 manufacturing and service work sites in Washington State	Convenience	QDASH SF-12 Symptoms severity	Hypothesis testing Concurrent validity Predictive validity
Fan (2011)	N= 465	Incident Cases: N=50 35.3 ± 10.2 (S) N=18 42.6 ±10.9 (C) Recovered Cases: N=46 35.5 ± 10.2 (S) N=41.9 ± 11.3 (C) Excluded Case: N=317 41.1 ± 10.7	Neck Or Upper Extremity Musculoskeletal Disorders (UEMSDs)	USA	Workplace walkthrough at 12 manufacturing and service work sites in Washington State	Convenience	QDASH SF-12 QDASH work module Severity	Responsiveness

Stock (2003)	Ontario N=119 Quebec N=93	Ontario: 39.7 ± 10.1 Quebec: 41.1 ± 10.0	Workers with neck and upper limb dysfunction	Canada	Workers from community private physiotherapy clinics	Convenience	NULI SIP SF-36	Internal consistency Reliability Structural validity Hypothesis testing Responsiveness
Lomond (2009)	N=32	N=16 40.1 ± 12.1 N=16 39.7 ± 13.2	Chronic neck and shoulder pain	Canada	Institutional rehabilitation programme, advertisement, research centre staff and social network	Convenience	SFA SPADI NDI NRS The Borg CR-10 scale	Test-retest, inter, intra-rater reliability Measurement error Hypothesis testing
Patekar (2010)	N=98	42.2 ± 7.85 (30-60)	Non-patient subjects with and without neck symptoms	UK	Institutional staff and students (institutional campus)	Convenience	SAMP	Hypothesis testing
Darne (2010)	N=95	44.53 ± 7.9 (30-60)	Non-patient subjects with and without neck symptoms	UK	Institutional staff and students (institutional campus)	Convenience	SAMP DASH	Hypothesis testing
Toulassidharane (2010)	N=190	41.8 ± 8.1 (30-59)	Non-patient subjects with and without neck symptoms	UK	Institutional staff and students (institutional campus)	Convenience	SAMP DASH	Hypothesis testing
Kulkarni (2010)	N=95	38.95 ± 7.22 (30-60)	Non-patient subjects with and without neck symptoms	UK	Institutional staff and students (institutional campus)	Convenience	SAMP DASH	Test-retest, inter, intra-rater reliability
Jain (2010)	N=95	44.5 ± 7.9 (30-60	Non-patient subjects with and without neck symptoms	UK	Institutional staff and students (institutional campus)	Convenience	SAMP DASH	Test-retest, inter, intra-rater reliability

Notes: DASH: Disability of Arm, Shoulder and Hand, QDASH: Quick Disability of Arm, Shoulder and Hand, SAMP: The Single Arm Military Press, NULI: The Neck and Upper Limb Index, (S): Symptomatic Cases, (C): Clinically Confirmed Cases, UK: United Kingdom, USA: United State of America.

Table 2.4: Methodological qualities of each study per measurement property

Study	Internal consistency	Test-retest, inter, intra-rater	Measurement error	Content validity	Structural validity	Hypothesis testing	Criterion Validity		Responsiveness
							Concurrent	Predictive	
DASH									
Huisstede (2009)						Poor			Poor
Mehta (2010) *						Poor	Good		
QDASH									
Fan (2008)						Poor	Poor	Poor	
Fan (2011)									Poor
Mehta (2010) *						Poor	Good		
NULI									
Stock (2003)	Fair	Fair		Excellent	Fair	Fair			Poor
SFA									
Lomond (2009)		Fair	Fair			Poor			
SAMP									
Patekar (2010)						Fair			
Darne (2010)						Poor			
Toulassidharane (2010)						Poor			
Kulkarni (2010)		Fair							
Jain (2010)		Fair							

*Notes: DASH: Disability of Arm, Shoulder and Hand, QDASH: Quick Disability of Arm, Shoulder and Hand, SAMP: The Single Arm Military Press, NULI: The Neck and Upper Limb Index, * This study is mentioned twice because of evaluating measurement properties of two instruments.*

Table 2.5: Quality of measurement properties per instrument for populations with NP

Instrument	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Criterion validity		Responsiveness	Practical properties		
							(C)	(P)		Precision	Acceptability	Feasibility
DASH	*na*	*na*	*na*	*na*	*na*	?	?	*na*	?	*na*	*na*	*na*
QDASH	*na*	*na*	*na*	*na*	*na*	?	?	?	?	*na*	*na*	*na*
NULI	+	+	*na*	*na*	+	+	*na*	*na*	?	*na*	*na*	*na*
SFA	*na*	+	+	*na*	*na*	?	*na*	*na*	*na*	*na*	*na*	*na*
SAMP test	*na*	++	*na*	*na*	*na*	+	*na*	*na*	*na*	*na*	*na*	*na*

(+++) or (---) strong evidence positive/negative results, (++) or (--) moderate evidence positive/negative results, (+) or (–) limited evidence positive/negative results, (±) conflicting evidence, (?) unknown, due to poor methodological quality, (na) no information available, (C) Concurrent, (P) Predictive.

2.3.4 Summary of measures

2.3.4.1 The Disability of Arm, Shoulder and Hand (DASH)

The DASH is a multidimensional PROM that was developed to evaluate the upper limb (hand, wrist, elbow and shoulder) disability and/or symptoms as a single functional unit (Hudak et al. 1996). The instrument uses 30-items related to difficulty when performing activity using the upper limb. The dimension physical function comprised 21-items, pain 5-items, emotional and social function 4-items. Each item is scored on a 1-5 scale. A total score is calculated by summing item scores and transforming them into a score from 0-100 where 0 equals no disability and 100 equals the most severe disability (Hudak et al., 1996).

Development and validation of the DASH

The development of the DASH was initially for the purpose of measuring ULD and was conducted in three phases (Hudak et al. 1996). First, the item generation phase, in which clinical experts and methodologists generated a list of 821 items after reviewing all the items included in 13 different questionnaires that were regularly used to address the health-related quality of life for patient populations with upper limb disorders. Second, the item reduction phase in which a set of 78-items was identified and field-tested in a cross-sectional study of 407 patients with different upper limb disorders in 20 centres in the USA, Canada and Australia. Equi-discriminative item total correlations supplemented with the patients rating of difficulty and importance were used to formulate the final 30-items DASH questionnaire. The measurement properties of the DASH were then extensively tested for a variety of upper limb (hand, wrist, elbow and shoulder) disorders and translated, culturally adapted, into over 40 languages (Westphal et al. 2002, Veehof et al. 2002, Soohoo et al. 2002, Offenbacher et al. 2003, Greenslade et al. 2004, Liang et al. 2004, Raven et al. 2008).

Subsequently, the DASH was validated to measure ULD in patients with NP in the following two studies. First, Huisstede et al. (2009) recruited 679 patients with NS-NP to investigate the validity and responsiveness of the DASH questionnaire. Participants were allocated into six sub-groups based on the location of their complaints and completed the DASH alongside the SF-12 (physical component summary scale (PCS) and mental component summary scale (MCS), severity of complaints, and persistence of complaints questionnaires at the baseline and at 6 months follow-up. Correlations were observed between the DASH and the other measures for all the sub-groups at the

baseline and 6 months follow-up. More than 75% of the hypotheses, which were formulated a priori for construct validity and responsiveness, were confirmed alongside an acceptable responsiveness ratio, and this suggested that the DASH questionnaire is valid and responsive measure of ULD for patients with NS-NP. Second, Mehta et al. (2010) investigated the validity of the DASH questionnaire in comparison with Quick Disability of Arm, Shoulder and Hand (QuickDASH), the Neck Disability Index (NDI) questionnaire, the Cervical Spine Outcome Questionnaire (CSOQ) and the Visual Analogue Scale (VAS) in patients with NP (n=66). The DASH showed high correlation and agreement with the QuickDASH alongside high correlation with the NDI, moderate correlations with the CSOQ and VAS, and this supported the validity of the DASH as a measure of ULD for patients with NP.

Measurement properties

Reliability: there were no studies investigating the reliability of the DASH in a population with neck pain. The construct validity of the DASH was assessed using Pearson correlations and the highest correlation, as expected and in the anticipated direction, was observed between the DASH score and the SF-12 (PCS) score at the baseline (range $r = 0.57$ and 0.63) when compared with the correlations between the DASH scores and the SF-12 (MCS) (range $r = 0.10$ and 0.33) and severity (range $r = 0.44$ and 0.55) (Huisstede et al. 2009). Construct validity was assessed again using the Pearson correlation and high correlation was observed between the DASH and the NDI ($r = 0.83$) and moderate correlation between the DASH and VAS ($r = 0.68$) (Mehta et al. 2010). The COSMIN 4-point checklist for construct validity was rated as 'poor' across these studies (see Table 2.4) because there was no information available on the measurement properties of the comparator OMs (see Appendix 3, Item 8 in Tables 1 and 2).

The measurement property criterion validity (concurrent) that met the COSMIN definition was assessed using the Bland and Altman plot to examine the level of agreement between the DASH and the QuickDASH in patients with NP. The mean differences between the DASH and QuickDASH alongside 2 standard deviation limits were 2.77 ± 10, which presented in a graph (Mehta et al. 2010). The COSMIN 4-point checklist for criterion validity was rated as 'good' (see Table 3.4) because of including moderate sample size (see Appendix 3, Item 3 in Table 2).

Responsiveness of the DASH questionnaire was assessed using the Guyatt's responsiveness ratio, which was over 1 for all the sub-groups (Huisstede et al., 2009). The COSMIN 4-point checklist for responsiveness was rated as 'poor' (see Table 3.4) because there was no information on the measurement properties of the comparator OMs, and the statistical methods applied (Guyatt's responsiveness ratio) were inappropriate for establishing responsiveness (see Appendix 3, Items12, 13 and 14 in Table 1).

2.3.4.2 The Quick Disability of Arm, Shoulder and Hand (QuickDASH)

Development and validation of the QuickDASH

The QuickDASH questionnaire is an 11-item PROM that was derived from the DASH questionnaire and designed to be a shorter measure of the disability and/or symptoms related to the upper limb (hand, wrist, elbow and shoulder) (Beaton et al. 2005). It is similar to the DASH, in that each item is scored on a 1-5 scale and the total score is derived by summing item scores and transforming them into a score from 0-100, where 0 equals no disability and 100 equals the most severe disability. The QuickDASH was also extensively tested for a variety of upper limb (hand, wrist, elbow and shoulder) disorders (Beaton et al. 2005, Gummesson et al. 2006, Matheson et al. 2006, Beaton et al. 2007, Mintken et al. 2009, Angst et al. 2009, Fayad et al. 2009, Gabel et al. 2009, Niekel et al. 2009, Poson et al. 2010, Angst et al. 2011, Franchignoni et al. 2011, Haas et al. 2011, Mardani-Kivi et al. 2013, Quatman-Yates et al. 2013, Nakamoto et al. 2014). The QuickDASH was recently updated to account for modern technology by replacing three items from its standard version with three other items related to the use of technology: (1) text or dial with your smart phone, (2) type on a keyboard and (3) use a computer mouse (Moradi et al. 2016).

The QuickDASH was also validated to measure ULD in patients/workers with NP in the following three studies. First, Mehta et al. (2010) investigated the construct and criterion validity of the QuickDASH against the DASH, NDI, CSOQ and VAS (n=66). The QuickDASH demonstrated high correlations and agreement with the DASH, high correlation with the NDI and moderate correlations with CSOQ and VAS, and this supports the validity of the QuickDASH for patients with NP (Mehta et al. 2010). Second, Fan et al. (2008) recruited 231 workers with a specific clinical diagnosis of neck or upper limb disorders alongside 175 workers with symptoms only (non-patient) to investigate the construct (discriminate) and criterion (concurrent and predictive)

validity of the QuickDASH in comparison with the SF-12 (PCS and MCS). The QuickDASH was administered by trained interviewers, in which participants rated their function capacity using the 11-item and the 4-item QuickDASH work questionnaire. The SF-12 was self-completed by all participants right after the interview. The QuickDASH demonstrated the ability to discriminate between workers in the two groups as well as between those with different symptom severity in the clinical cases group. A moderate correlation was observed between the Quick DASH and SF-12 (PCS) on workers with neck or upper limb disorder (clinical cases), and this supports the use of the QuickDASH as a valid measure of ULD for workers with NP (Fan et al. 2008). Third, Fan et al. (2011) investigated the responsiveness of the QuickDASH as well as the SF-12 (PCS and MCS) to change in active workers with neck or upper limb disorders (clinical cases) alongside symptomatic and non-symptomatic workers in a one-year follow-up. The standard QuickDASH and the work module demonstrated the ability to detect change for all the sub-groups of workers, which support the responsiveness of the QuickDASH for workers with NP (Fan et al. 2011).

Measurement properties

No studies investigated the reliability of the QuickDASH in a NP population. Construct validity was evaluated using Pearson correlation, and high correlation was observed between the QuickDASH and the NDI ($r = 0.82$), and moderate correlations were observed between the QuickDASH and the CSOQ components (neck pain, shoulder and arm pain, physical symptoms, functional disability and psychological distress) and VAS ($r = 0.65$, 0.57, 0.68, 0.59, 0.58 and 0.64) respectively. These relationships were statistically significant ($p < 0.01$) (Mehta et al. 2010). Construct validity was assessed against the SF-12 using the two-factor analysis of variance (ANOVA), in which the standard QuickDASH and the work module scores were higher when compared with the SF-12 (PCS) in the clinical cases with neck or upper limb disorders (Fan et al. 2008). The COSMIN 4-point checklist for construct validity was rated as 'poor' in both studies (see Table 3.4) because there was no information available on the measurement properties of the comparator OMs in the first study, it was unclear what was expected and the statistical methods applied were inappropriate in the second study (see Appendix 3, Items 8 in Table 2 and item 4 in Table 3).

Criterion validity (concurrent) was evaluated using the Bland and Altman plot to examine the agreement between the QuickDASH and the DASH in patients with NP.

The mean differences between the DASH AND QuickDASH alongside 2 standard deviation limits were 2.77 ± 10, which presented in a graph (Mehta et al. 2010). Criterion validity (concurrent and predictive) was assessed using Spearman rank correlation between QuickDASH and the SF-12 for concurrent validity and the odds ratios for predictive validity (Fan et al. 2008). The COSMIN 4-point checklist for criterion validity was rated as 'good' in the first study because of the sample size and rated as 'poor' in the second study because the criterion used cannot be considered as a reasonable gold standard (see Table 3.4), (see Appendix 3, Item 3 in Table 2 and Item 4 in Table 3).

Responsiveness of the QuickDASH was assessed using the Effect Size (ES) and the Standard Response Mean (SRM) for the QuickDASH and SF-12. The ES and the SRM for the QuickDASH were > 0.08, and this indicates large change between the workers in all groups but one sub-group (self-reported symptomatic case) had moderate ES and SRM. Meanwhile, the scores for the SF-12 (PCS) decreased as expected (Fan et al. 2011). The COSMIN 4-point checklist for responsiveness was rated as 'poor' (see Table 3.4) because the statistical methods applied were inappropriate measures of responsiveness (see Appendix 3, Items 14 in Table 4).

2.3.4.3 The Neck and Upper Limb Index (NULI)

The NULI questionnaire is a short English and French language multidimensional Patient-Reported Outcome Measurement (PROM) which was developed to measure the functional status for workers with NP and ULD (Stock et al. 2003). The questionnaire uses a 20-item index to evaluate the impact of neck and upper limb disorders on physical function, work, psychosocial limitations and sleep. The dimension physical function/physical activity dimension comprised 7 items, work comprised 4 items, psychosocial comprised 6 items, sleep comprised 2 items and 1 item related to the iatrogenic effect of assessment and treatment (Stock et al. 2003). Section A of the questionnaire, questions 1-11 are scored on a 1-7 scale, where 1 equals no difficulties at all and 7 equals cannot do. Section B, questions 12-20 are scored on a 1-7 scale where 1 equals never and 7 equals all the time (stock et al. 2003).

Development and validation of the NULI

The development of the NULI started with item generation, in which a comprehensive review of all the relevant scientific literature was carried out to identify all available

outcome measures. This stage involved the participation of researchers and clinicians who were working with patients with NP and ULD. This led to the generation of 175items which fell into 12-dimensions and which were thought to be affected by the NP and ULD.

The item reduction stages involved interviewing workers with NP and ULD (n=33) as well as surveying clinicians (physiotherapists, occupational therapists, rheumatologists, physiatrist, orthopaedic surgeons and doctors managing patients with NP and ULD) (n=30). Participating patients and clinicians were requested to identify and list all activities that were frequently affected by the NP and ULD. This led to the formulation of the 20-item NULI questionnaire.

The NULI questionnaire was developed and validated in a single study. Stock et al. (2003) recruited 119 English-speaking and 93 French-speaking workers from eight different private physiotherapy clinics in Quebec (Canada) to investigate the reliability, validity and responsiveness of the NULI. The NULI demonstrated strong reliability (test-retest), good content, convergent and discriminate validity as well as responsiveness. This provides support for the use of the NULI questionnaire for workers with NP and ULD.

Measurement properties

Reliability (internal consistency) was evaluated, in which Cronbach alpha was calculated for the English-speaking participants (0.90), French-speaking participants (0.92) and the final 20-item NULI (0.93). Factor analysis was used to evaluate the dimensionality of the NULI, in which the first factor correlated between 0.74 and 0.87 with the 4-item work construct, the second factor correlated between 0.47 and 0.78 with the 6-item construct about physical activity (excluding the question about activity of leisure), the third factor correlated between 0.32 and 0.87 with the 6-item psychosocial construct and the fourth factor correlated between 0.87 and 0.82 with the 2-item sleep construct (Stock et al. 2003). The COSMIN 4-point checklist for internal consistency was rated as ‘fair’ in this study (see Table 3.4), and this was because it was unclear how missing items were handled (see Appendix 3, Item 3 in Table 5). The test-retest and inter-rater reliability were assessed using the Interclass Correlation Coefficient (ICC). For English-speaking subjects ICC = 0.88, French-speaking ICC = 0.85 and the final NULI ICC = 0.83 (Stock et al. 2003). The COSMIN 4-point checklist for this

measurement property was rated as 'fair' in this study (see Table 3.4) because it was unclear how missing items were handled and it was also unclear if patients were stable in the interim period on the construct to be measured (see Appendix 3, Items 2 and 7 in Table 5).

Content validity of the NULI was examined by asking participants (workers with NP and ULD) and clinicians (physiotherapists, occupational therapists, rheumatologists, physiatrists, orthopaedic surgeons and doctors treating patients with NP and ULD) to identify items related to function which were affected by NP and ULD (Stock et al. 2003). The COSMIN 4-point checklist for content validity was rated as 'excellent' in this study (see Table 3.4) as all items were adequate (see Appendix 3, Items 1, 2, 3, 4 and 5 in Table 5).

The structural validity of the NULI was assessed using confirmatory factor analysis and correlation was observed between the four items about work and the first factor (range 0.74 and 0.87). The six items about physical activity (excluding the question about leisure activity) correlated with the second factor (range 0.47 and 0.78). Correlations were also observed between the six items about psychological effect and the third factor (range 0.32 and 0.87) and finally between the two items about sleep and the fourth factor (range 0.87 and 0.82) (Stock et al. 2003). The COSMIN 4-point checklist for structural validity was rated as 'fair' in this study (see Table 3.4) because it was unclear how missing items were handled and it was unclear if patients were stable between measurements (see Appendix 3, Items 3 and 7 in Table 6).

Convergent validity of the NULI was evaluated using Pearson correlation analysis and revealed significant moderate correlations between the NULI, the Sickness Impact Profile (SIP) ($r = 0.66$ $P < 0.001$) and the SF-36 ($r = 0.50$ $p< 0.001$) (Stock et al. 2003). The COSMIN 4-point checklist for convergent validity was rated as 'fair' in this study (see Table 3.4) because it was no adequate description for the measurement properties of the comparator OMs (see Appendix 3, Item 8 in Table 5).

Responsiveness of the NULI was examined in this study using the Standard Response Mean (SRM) that was 1.48 (95% CI:1.1 – 1.8) for English-speaking subjects and 1.63 (95%CI: 1.3 – 2.0) for French-speaking subjects (Stock et al. 2003). The COSMIN 4point checklist for responsiveness was rated as 'poor' (see Table 3.4) due to the

statistical methods applied being inappropriate measures of responsiveness (see Appendix 3, Item 14 in Table 5).

2.3.4.4 The Shoulder Functional Assessment (SFA)

The SFA is a clinician-reported (objective) OM designed to quantify the upper limb functional capacity in workers with neck and shoulder pain. The SFA protocol consists of tasks involving shoulder range of motion (ROM) in both flexion and abduction and cumulative power output (PO) accumulated over 10 seconds during a repetitive arm pushing/pulling task on a horizontal plane at shoulder level. These tasks were included since they represent movements that are often impaired in people with chronic neck and shoulder pain (Donovan and Paulos 1995, Hoozemans et al. 2002). The tasks were assessed before and after performing a repetitive arm task until scoring 8 on the Borg CR10 scale or an 11-point numeric pain rating scale. The scoring system is determined by the average repetitive task duration with expectation that the shorter duration represents poor performance and therefore a higher level of NP and ULD (Lomond and Cote 2009).

Development and validation of the SFA

The SFA protocol is conducted using the Baltimore Therapeutic Equipment Work Simulator II (Sim II) (BTE-Tech©, Baltimore, MD), which is a functional capacity evaluation tool commonly used to measure upper limb functional capacity (Bhambhani et al. 1993, Trossman et al. 1990). Sim II was introduced in 1979 to be used as an assessment instrument and therapy tool for workers with ULD (Coleman et al. 1996). Sim II consists of a software-based controller interface, a position adjustable exercise head with resistance control, and a set of interchangeable attachments to simulate a variety of work tasks.

The SFA was validated in a single study. Lomond and Cote (2009) recruited 16 workers with chronic neck/shoulder pain and matched 16 healthy subjects to investigate the test-retest reliability and measurement error, as well as the discriminant validity of the SFA in workers with neck/shoulder pain and healthy subjects. There was no significant difference in the cumulative power output (PO) task between groups or testing sessions, suggesting that workers with neck/shoulder pain were able to perform that task as effectively as the healthy subjects. This indicates poor discriminant validity of the SFA

in relation to the PO task. The SFA showed good to excellent reliability in both groups (Lomond and Cote 2009).

Measurement properties

Reliability (test-retest) of the SFA was examined using the ICC. For the flexion task ICC was 0.95 (control group) and 0.92 (pain group), for abduction ICC was 0.85 (control group) and 0.87 (pain group) and for cumulative power output ICC was 0.94 (control group) and 0.53 (pain group) (Lomond and Cote 2009). The measurement error for the SFA was assessed using the Standard Error of Measurement (SEM) and the Minimal Detectable Change (MDC). For the flexion task, SEM was 4.72 (control group) and 14.76 (pain group), abduction SEM was 6.06 (control group) and 24.35 (pain group), cumulative power output SEM was 7.52 (control group) and 30.25 (pain group). For flexion, MDC was 11.01 (control group) and 34.44 (pain group), abduction MDC was 14.15 (control group) and 56.81 (pain group), and cumulative power output MDC was 17.54 (control group) and 70.59 (pain group) (Lomond and Cote 2009). The COSMIN 4-point checklist for reliability and measurement error were rated as 'fair' in this study (see Table 3.4) due to moderate sample size being used and it was unclear whether participants were stable in the interim period on the construct to be measured (see Appendix 3, Items 3 and 7 in Table 6).

Construct validity of the SFA was assessed by evaluating the SFA's sensitivity to discriminate between workers with neck/shoulder pain and healthy subjects (Lomond and Cote 2009). The COSMIN 4-point checklist for construct validity was rated as 'poor' in this study (see Table 3.4) as there were no hypotheses formulated a priori regarding the correlations or differences and it was unclear what was expected (see Appendix 3, Item 4 in Table 6).

2.3.4.5 The Single Arm Military Press (SAMP) test

The SAMP test is a performance-based test developed specifically to evaluate the upper limb functional capacity in populations with NS-NP (McLean et al., 2010a). The SAMP test, which is a strength and endurance-based test, involves repeatedly lifting a 3kg hand-weight overhead from the shoulder level for 30 seconds. The test score is the number of repetitions correctly completed within the 30 seconds, with higher scores representing a lower level of NS-NP and ULD (McLean et al., 2010a).

Development and validation of the SAMP test

The SAMP test was developed following findings which showed that there was a strong association between NP and ULD, with patients reporting the greatest difficulties with heavy household chores, gardening, carrying heavy objects and overhead activities (McLean 2007, McLean et al. 2010). In addition, there was no performance-based measure available to evaluate the ULD when assessing patients with NP and the DASH questionnaire, which was poorly validated for NP patients, but was the only available OM at that time (McLean 2007).

The SAMP test was validated in a series of preliminary unpublished studies. Kulkarni (2010) recruited 95 female participants with/without neck symptoms to investigate the reliability (inter-and intra-rater) of the SAMP test. Almost perfect inter-and intra-rater reliability and agreement were found in both symptomatic and asymptomatic participants. This indicates that the SAMP test is a reliable measure of ULD (Kulkarni 2010). Jain (2010) also recruited a cohort of 95 female participants with/without neck symptoms to investigate the reliability (inter-and intra-rater) of the SAMP test. Similar to the previous study, almost perfect reliability was found for both symptomatic and asymptomatic participants, which, again, supports the reliability of the SAMP test as a measure of ULD (Jain 2010). Patekar's (2010) study recruited 98 female subjects with/without neck symptoms to investigate the construct (discriminate) validity of the SAMP test versus the DASH. Symptomatic and asymptomatic female participants performed the SAMP test using their dominant hand with a 3-kg hand weight. Correlation was observed and the asymptomatic participants were significantly better than those symptomatic participants in the SAMP performance, and this indicates that the SAMP test has the capacity to discriminate between symptomatic and asymptomatic populations (Patekar 2010). A study by Darne (2010) recruited 95 female subjects with/without neck symptoms to investigate the construct (convergent) validity of the SAMP test. Participants in this study completed the DASH questionnaire and performed the SAMP test in a single session. A highly significant negative correlation, as expected, was observed between the SAMP performance and the DASH score in the anticipated direction. This indicates that the SAMP test is a valid measure of ULD (Darne 2010). Finally, Toulassidharane (2010) recruited a cohort of 95 female subjects with/without neck symptoms to investigate the construct (convergent) validity of the SAMP test. Participants in this study also completed the DASH questionnaire before performing the SAMP test. A highly significant negative correlation, as expected, was

also observed between the SAMP performance and the DASH score, which supports the validity of the SAMP test as a measure of ULD (Toulassidharane 2010).

Measurement properties

Reliability was assessed using the ICCs; this revealed a high level of inter-rater reliability, with a 95% Confidence Interval (CI) (ICC = 0.99), while for intra-rater reliability (ICC = 0.94) in a symptomatic population of women (Kulkarni 2010). Reliability was assessed again using the ICCs and also revealed a high level of interrater reliability, with a 95% CI (ICC = 0.98), while for intra-rater reliability (ICC = 0.99) in an asymptomatic population of women (Jain 2010). The COSMIN 4-point checklist for reliability was rated as 'fair' in the two studies (see Table 3.4) as it was unclear if patients were stable in the interim period (see Appendix 3, Item 7 in Tables 7 and 8).

Construct (discriminate) validity was assessed using the unpaired t-test, which revealed that asymptomatic female participants significantly and substantially out-performed those symptomatic participants (mean SAMP scores asymptomatic participants=30, mean SAMP scores symptomatic participants=18, $p<0.001$, n=98) (Patekar 2010). Construct (convergent) validity was assessed using Pearson correlation and highly significant negative correlation was observed between the SAMP performance and the DASH score ($r = -0.800$, $p<0.001$, n=95) (Darne 2010). Construct (convergent) validity was assessed, again using Pearson correlation, which also revealed a highly significant negative correlation between the SAMP performance and the DASH score ($r = -0.814$, $p<0.001$, n=95) (Toulassidharane 2010). The COSMIN 4-point checklist for construct validity was rated as 'fair' in the first study because hypotheses were not formulated a priori, but it is possible to deduce what was expected (see Appendix 3, Item 4 in Table 9). Construct validity was rated as 'poor' in the other two studies because there was no information on the measurement properties of the comparator OMs (see Appendix 3, Item 8 in Tables 10 and 11).

2.4 Discussion

2.4.1 Summary and discussion of the main findings

The overall aim of this review was to identify, summarise and critically examine all available studies on the measurement properties of all available OMs that have been developed or validated to measure ULD for patient populations with NP. This was done in order to make recommendations regarding relevant OMs and provide the background

and justification for future research needs. In total 11 studies evaluating the measurement properties of five clearly defined and reproducible OMs were included within this review. To recommend OMs that can enable clinicians to identify and quantify ULD when assessing and managing their patients with NP, it is essential that sufficient information is available regarding the development process of this OM and the measurement model, if relevant, alongside adequate evaluation of all relevant measurement and practical properties in the target population. However, synthesis of the results demonstrated a paucity of high-quality evidence and significant methodological and quality issues, as well as missing evidence for relevant and essential measurement and practical properties. This prevented a clear recommendation for any of the five included OMs. Evidence for the five identified and included OMs was limited, unknown or unavailable.

Critical appraisal of all the included OMs using the COSMIN 4-point checklist demonstrated that other than content validity for the NULI questionnaire and concurrent validity for the DASH/QuickDASH; none of the identified and reviewed OMs reported an 'excellent or good' rating in its measurement properties. The measurement model, which indicates whether an OM is based on a formative or reflective model, was not described in the three identified questionnaires (DASH, QuickDASH and NULI). Information regarding the measurement model has implications when evaluating the measurement properties of a PROM as internal consistency and structural validity are relevant for evaluation only if the OM is based on a reflective model (Mokkink et al. 2010a). Other than the NULI, evidence for face/content validity and the practical properties including acceptability and feasibility was not identified for any of the reviewed OMs. Acceptability and feasibility of OMs for patients and clinicians are necessary for the clinical utility of an OM (Fitzpatrick et al. 1998). There was no evidence of patient involvement as a research partner in the development/evaluation of any of the OMs, with the exception of the NULI. Patient involvement as a research partner is considered essential to ensure the relevance, comprehensiveness and validity of patient-centred outcome assessment (Mayer 2012, Staniszewska et al. 2012). One clinician-reported (objective) measure (SFA) was developed for workers only and involves the use of expensive equipment. This is likely to limit its use in clinical practice (Fitzpatrick et al. 1998). Since brevity is crucial in clinical practice, the QuickDASH, NULI and the SAMP test are promising clinical measures if adequately

validated. However, the SAMP test is the only performance-based measure that was developed specifically to measure ULD for patient populations with NP.

2.4.2 Evaluation of identified outcome measures

There is substantial evidence that the DASH and QuickDASH well performing measures in patients with upper limb (hand, wrist, elbow and shoulder) disability/symptoms only (Huang et al. 2015, Kennedy et al. 2013). Limited evidence indicated that the NULI and SFA are reliable and valid measures for workers with neck and upper limb disorders (Lomond and Cote 2009, Stock et al. 2003).

2.4.2.1 Development

The development of the DASH, QuickDASH and NULI questionnaires incorporated the same techniques, which included item generation and item reduction before formulating and testing the final version. Sufficient information was available for each stage and both patients and clinicians participated in the item generation and in the item reduction stages. This demonstrates excellent acceptability, feasibility and face/content validity for these measures. However, the development of the DASH and QuickDASH did not involve patients with NP (the target population), and the development of the NULI included Canadian workers only (English/French-speaking). The SFA and the SAMP test do not provide sufficient information on their respective development processes. Application of an instrument for a purpose and/or population other than which it was intended coupled with inadequate levels of validation, limits meaningful interpretation and recommendations that can be made about the clinical utility of these measures (Fitzpatrick et al. 1998, de Vet et al. 2011).

2.4.2.2 Reliability

Internal consistency was reported only for the NULI questionnaire and rated as ‘fair’. This rating was influenced by a lack of information on the percentage of missing items or how missing items were handled. A large number of missing items in a questionnaire can introduce bias in the results and have a significant impact on the OM’s scores (Mokkink et al. 2010a). Reliability (test-retest, inter- and intra-rater) was not assessed for the DASH, QuickDASH in the target population (patients with NP). The NULI questionnaire, SFA and the SAMP test each received a ‘fair’ rating for reliability. It has been suggested that patients involved in a reliability study should be stable in the period between administrations on the construct being measured (de Vet et al. 2011). This is to

facilitate the decision regarding consistency of scores for the OM under study. This was not the case for the NULI, SFA and the SAMP test and it was unclear whether patients were stable between administrations. In addition, the moderate sample size used in the SFA affected the rating for reliability. Measurement error, which is an extension of the reliability domain, often mirrored the results of the reliability (Mokkink et al. 2010a). Measurement error was assessed for the SFA only and rated as 'fair'. Similar to reliability, moderate sample size affected the rating of measurement error. In a reliability study, adequate sample size is crucial to obtaining an acceptable confidence interval (CI) around the calculated Interclass Correlation Coefficient (ICC) when evaluating reliability and/or measurement error (de Vet et al. 2011).

2.4.2.3 Validity

Face/content validity was not reported for the DASH, QuickDASH, SFA or the SAMP test. The NULI questionnaire evaluated face/content validity, which was the only measurement property that received an 'excellent' rating in this review. This rating was influenced by the adequate judgement of relevance and the comprehensiveness/coverage of items using experts (researcher/clinicians) as well as the targeted populations in the development stages (Mokkink et al. 2010a). Structural validity is part of the measurement property, construct validity. It is relevant for evaluation only if the OM under study is a PROM and based on a reflective model (de Vet et al. 2011). Structural validity was reported only for the NULI questionnaire and was rated as 'fair'. This was affected by a lack of information about missing items and how missing items were handled. A large number of missing items can introduce bias in the results and have a significant impact on the scores of the OM (Mokkink et al. 2010a). Hypothesis testing was the only measurement property which was investigated in all five included OMs. Hypothesis testing was rated as 'fair' and/or 'poor' in all the identified OMs. This rating was affected mostly by the lack of adequate description of the measurement properties on the comparator OMs. In convergent validity, to determine whether negative results are due to poor validity of the OM under study; the comparator OMs should be appropriately described regarding their construct and adequately validated in the target population (Mokkink et al. 2010a). Criterion validity was evaluated for the DASH and QuickDASH and rated as 'poor' because the criterion cannot be considered an adequate gold standard. Using criterion validity to assess the validity of an OM primarily requires identifying criterion that are an adequate gold standard and a priori formulation of a hypothesis that the OM under study is as good as

the gold standard (de Vet et al. 2011). Gold standard criterion is defined here as the OM which represents the true state of the construct of interest (de Vet et al. 2011). However, there is no gold standard which exists for health-related PROMs because they are subjective tools, which often measure patients' perceptions and opinions about their pain and/or disability (Mokkink et al. 2010a). Consequently, with the exception of the QuickDASH, with the DASH as a gold standard comparator, there is no expectation that any of these ULD OMs should have criterion validity.

2.4.2.4 Responsiveness

Responsiveness is considered a longitudinal validity, which indicates that the OM under study should be administered at least twice during a longitudinal study. The comparator OM, preferably a gold standard, should be administered at the same time to confirm the change score in the construct of interest. In addition, applied statistical methods should be appropriate to determine the validity of the change score. This was not the case for all studies reporting responsiveness in this review. The DASH, QuickDASH and NULI questionnaires were evaluated for responsiveness and received a 'poor' rating. This rating was primarily influenced by a lack of reporting description regarding the validation of the comparator OMs coupled with the application of inappropriate statistical methods, such as the Guyatt's responsiveness ratio and the Effect Size (ES). Guyatt's responsiveness ratio calculates the Minimal Important Change (MIC) on the OM under investigation, divided by the Standard Deviation (SD) of the change score (Guyatt et al. 1987). This provides information on the interpretability/meaning of change score rather than the validity of change score (de Vet et. 2011). ES calculates the mean change score divided by the SD and measures the magnitude of change score rather than the validity of change score (de Vet et al. 2011).

2.4.3 Strengths of the review

The major strengths of this review include the comprehensive search strategies developed; the wide scope of the search, including grey literature and contacting authors and/or developers of specific measures for other published or unpublished studies; the independent appraisal of the methodological quality of the included studies and the data extraction; use of the COSMIN checklist (Mokkink et al. 2010a, Terwee et al. 2012) as well as the reporting of the review in accordance with the PRISMA statement (Liberti et al. 2009, Moher et al. 2009). In addition, this is the first systematic review that has

sought to identify and evaluate the measurement properties of all available ULD OMs developed or validated for use in patients with NP.

2.4.4 Limitations of the review

Although the search strategies in this review were limited to English-language publications, English-language abstracts for non-English publications were reviewed and one study only was excluded, and this was due to irrelevance not language. This suggests that the likelihood of selection bias is low. The level of evidence criteria in Table 3.2 that was suggested by the Cochrane Back Review Group (van Tulder et al. 2003, Furlan et al. 2009) was originally proposed for systematic reviews conducted on clinical trials. However, it has been used in similar studies and found to be applicable to reviews investigating the measurement properties of health-related OMs (Schellingerhout et al. 2011, Schellingerhout et al. 2012).

2.4.5 Clinical implications

Clinically, it is recognised that there is a strong relationship between NP and ULD and the presence of NP is a risk factor for the development of ULD (Walker-Bone et al. 2004, Frank et al. 2005, Bot et al. 2005, Huisstede et al. 2006, Rasmussen et al. 2008, Feleus et al. 2008). In addition, co-existing shoulder problems may lead to NP becoming recurrent, persistent or disabling (Eriksen et al. 1999, Bot et al. 2005, McLean et al., 2010a). Routine utilisation of standardised ULD OMs in the assessment and during the management of patients with NP is essential since it may play an integral part in influencing clinical outcomes (McLean et al. 2011, Osborn and Jull 2013). The findings from this review highlighted that promising ULD OMs exist; however, they have significant methodological and quality issues as well as missing evidence for relevant and very important measurement and practical properties. Consequently, at this time, none of these measures can be formally recommended for use in a clinical context.

The SAMP test was the only identified performance-based measure that was developed specifically to identify and quantify ULD in the assessment of patient populations with NP and to monitor its progress during rehabilitation (McLean et al. 2010a). It is a physical performance test that requires a patient to use multiple joints to perform a task, which represents some construct of function including strength and endurance. Therefore, it has a greater likelihood of accurately measuring the upper limb functional capacity more than PROMs (Curb et al. 2006, Pinheiro et al. 2016). The SAMP test also

has the advantage of being able to be efficiently administered by individual/clinicians of varying experience, in any setting, using minimal equipment within less than 2-minutes. The SAMP test therefore, is convenient, efficient and inexpensive. However, the SAMP test has been validated as a 3-kg hand weight for non-patient populations with NS-NP and healthy subjects. Given that patients with NS-NP and ULD are likely to experience greater severity of their symptoms (McLean et al. 2011, Osborn and Jull 2013), a 3-kg hand weight is perhaps too heavy for their use. Therefore, at this time, the 3-kg SAMP test, in spite of its potential utility, cannot be recommended for use in a patient population with NP and ULD.

2.4.6 Research implications

Further adequate validation is required for at least one of the three promising measures (QuickDASH, NULI and the SAMP test) incorporating the COSMIN recommendations. The SAMP test was selected for further validation in this thesis. It requires adequate validation, which should investigate the acceptability and feasibility of the weight used in its practical application on populations with NP and ULD. This is to improve its clinical utility for this patient group. Furthermore, important measurement properties such as the reliability and validity of the SAMP need to be explored in the same patient group. These measurement and practical properties will be further investigated in chapters 5 and 6.

2.5 Conclusion

In the absence of high-quality studies and inadequate reporting of essential measurement and practical properties, application of the identified ULD OMs cannot be recommended in populations with NP until acceptable evidence is established. Further research should incorporate COSMIN recommendation during the design of developmental or evaluative studies of these OMs. The involvement of key stakeholders, including patients and clinicians is essential to ensure that the OM is relevant, acceptable and feasible.

The next chapter is a review of the literature relating the current evidence-based management practice within the scope of physiotherapy for patients with NS-NP. It considers a wide range of treatment approaches that are routinely available to physiotherapists.

Chapter 3: Evidence of the currently recommended treatment approaches for the management of non-specific neck pain: A literature review.

3.1 Introduction

This chapter describes a narrative literature review that explored current evidence-based management practice within the scope of physiotherapy for patients with NS-NP. This was based primarily on recent evidence-based guidelines and systematic reviews. The findings were used to inform the development of the subsequent UK national survey of neck pain reported in chapter 4.

Various physiotherapy treatment approaches have been described and tested for the management of patients with NS-NP. The general classification of these treatment approaches can be divided into "active" modalities (that is, an active physical movement in which the patients engage muscles of an injured part of their body to create that movement), where patients also take control of their rehabilitation under the supervision of physiotherapists (Capersen et al. 1985), and this includes all the types of exercise, patient education programmes and the McKenzie method. Passive modalities, in which the patient depends on an external stimulus (the physiotherapist) to apply movement or treatment to an injured part of their body, and this includes manual therapy, electrotherapy, massage therapy acupuncture and traction (McLean and Moffett 2006).

The most frequently used treatment approaches have frequently been examined for their effectiveness when used in isolation. However, they have also been increasingly investigated when used as part of a multimodal treatment approach, which combines at least two treatment approaches with exercise being a key component of any combination (Gross et al. 2007, Hurwitz et al. 2008). Multimodal approaches to management of NS-NP are categorised in this thesis into three common different packages: (1) exercise, manual therapy and any form of patient education, (2) exercise and manual therapy and (3) exercise and any other treatment approach.

3.1.1 Aim

This comprehensive literature review has scoped the most recent high-quality evidence of the recommended physiotherapy approaches, which are frequently used for managing patients with NS-NP. This aim was addressed by meeting the following objectives:

1. To determine current recommended non-invasive treatment approaches for NS-NP.
2. To determine the effectiveness of these approaches when used in isolation or when used as part of a multimodal approach.

3.2 Methods

3.2.1 Data source

The following databases were searched to identify all relevant systematic reviews, randomised controlled trials (RCTs), or controlled clinical trials (CCTs) related to neck function and treatment approaches: MEDLINE (EBSCO), CINAHL (EBSCO), SPORTDiscus (EBSCO), PsycINFO (ProQuest), PubMed (NLM), AMED (OvidSP), the Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment Database, NHS Economic Evaluation Database, and Cochrane Collaboration) (Wiley), Web of Science (Thomson Reuters), and Google Scholar (Google). A RefWork database was used to manage all references.

3.2.2 Search strategy

A search strategy combining title/abstract words and database subject headings, where available, relating to neck function and treatment approaches was used to capture all relevant systematic reviews, RCTs, or CCTs. The search comprised two facets: (1) terms relating to neck function and (2) terms relating to treatment approaches. To capture the phrase “neck function”, title/abstract words, all possible synonyms and database subject headings (e.g. MeSH) for neck were combined with those for function using the Boolean operators AND and OR, alongside truncation, phrase searching and proximity operators. The search strategy for MEDLINE (EBSCO) is as follows. This search syntax detailed below were adapted for use on other information resources used in the search.

Box 3.1: Search strategy

Function terms: words in title and abstract: function*, pain*, dysfunction*, disability*, disorder*, problem*, disc*, diseas*, "neck" n5 Function* ti,ab. "neck" n5 pain*. ti.ab. "neck" n5 dydfunction*. ti.ab. "neck" n5 disabilit*. ti,ab. "neck" n5 disorder*. ti,ab. "neck" n5 problem*. ti,ab. "neck" n5 disc*. ti,ab. OR "neck" n5 diseas*. To restrict the search to the context of treatment approaches, the above searches using the conjunction AND was combined with the following search terms/synonymous and database subjects headings: massage*, "cognitive behavioural therapy" "CPT", electrotherapy*, traction*, collar*, advice* "back school*", education*, "manual therapy", mobilization*, manipulation*, exercise*, stretching*, strengthen*, OR "physical modalit*.

Note:(ti) = title field, (ab) = abstract field, (/) = MeSH, asterisk (*) denotes any character, ("") = phrase search, (n5) = adjacency within five words.

3.2.3 Study selection

All published systematic reviews, RCTs, or CCTs, in the English language from 1995 onwards were considered for inclusion if: (1) the sample involved an adult population with NS-NP, (2) involved a physiotherapeutic intervention, which is defined here as any non-invasive, non-surgical and non-pharmacological treatment used for the management of NP. Studies that involved participants with any serious pathology, systemic disease, neurological deficit, major trauma, any inflammatory condition, or incorporated non-physiotherapeutic intervention were excluded.

3.3 Findings

A total of 31 systematic reviews and 8 RCTs were located and informed the findings set out below. A table of included studies indicating the modalities investigated is shown below in Table 3.1.

The findings as follows are divided into four sections. First, passive treatment approaches including manual therapy (manipulation and mobilisation), massage therapy, electrotherapy, acupuncture and traction. Second, active treatment approaches including therapeutic exercise, the McKenzie method and Feldenkrais. Third, patient education. Finally, the multimodal management approach packages are discussed.

Table 3.1: The modalities investigated

Name	Study Type	Manual Therapy	Massage Therapy	Electro-therapy	Acupuncture	Traction	Exercise Therapy	MDT	Feldenkrais	PE	Multimodal Approach
Bearman and Shafarman 1999	RCT								√		
Lundblad et al. 1999	RCT								√		
Gross et al. 2002	SR	√									√
Sarig-Bahat 2003	SR						√				
Gross et al. 2004	SR	√									√
Clare et al. 2004	SR							√			
Vernon et al. 2005	SR	√									
Kay et al. 2005	SR						√				√
Sarigiovannis and Hollins 2005	SR	√									
Kroeling et al. 2005	SR			√							
Gemmell and Miller 2006	SR	√									
Haraldsson et al. 2006	SR		√								
Trinh et al. 2006	SR				√						
Vas et al. 2006	RCT				√						
Graham et al. 2006	SR					√					
Vernon and Humphreys 2007	SR	√									
Vernon et al. 2007	SR	√	√							√	
Ezzo et al. 2007	SR		√								
Macaulay et al. 2007	SR	√									

Itoh et al. 2007	RCT				√						
Gross et al. 2007	SR		√	√	√	√	√			√	√
O'Leary et a. 2007	SR						√				
Ylinen et al. 2007	SR						√				
Graham et al. 2008	SR					√					
Hurwitz et al. 2008	SR			√			√			√	√
Haines et al. 2008	SR									√	
Bernaards et al. 2008	RCT									√	
Hakkinen et al. 2008	SR						√				
Graham et al. 2008	SR					√					
Fu et al. 2009	SR				√						
Gross et al. 2010	SR	√		√	√						
Miller et al. 2010	SR										√
Liang et al. 2011	RCT				√						
Kay et al. 2012	SR						√				√
Bertozzi et al. 2013	SR						√				
Kroeling et al. 2013	SR			√							
O'Riordan et al. 2014	SR						√				
Perez et al. 2014	RCT	√									
Sherman et al. 2014	RCT		√								

SR: Systematic Review, RCT: Randomised Controlled Trial, MDT: Mechanical Diagnostic and Therapy, PE: Patient Education.

3.3.1 Passive treatment approaches

3.3.1.1 Manual therapy (manipulation and mobilisation)

Manual therapy (manipulation and mobilisation) is a commonly used intervention either alone or combined with different therapy modalities for patients with NS-NP (Gross et al. 2002). Cervical manipulation is a passive technique applied to a joint in the cervical spine by incorporating localised high velocity with low amplitude thrust to reduce pain, improve function and restore optimal joint range of motion (Herzog 2010). Meanwhile, cervical mobilisation is any manual therapy technique that incorporates a sequence of passive movement with varying speed and amplitude to a joint in the cervical spine for the purposes of reducing pain and increasing range of motion (Gross et al. 2004).

Manual therapy was investigated in 9 systematic reviews and one recently published trial of a non-invasive intervention for NS-NP. The effectiveness of manipulation or mobilisation remains inconclusive when used as a single intervention for the management of patients with NS-NP (Gross et al. 2004, Vernon et al. 2005, Sarigiovannis and Hollins 2005, Gemmell and Miller 2006, Vernon and Humphreys 2007, Vernon et al. 2007, Macaulay et al. 2007). In their updated review, Gross et al. (2010) identified 27 trials (n=1522) that assessed the effectiveness of manipulation and mobilisation when compared with no treatment, a sham treatment or another intervention such as medication, acupuncture, heat, electrotherapy and massage. Moderate quality evidence suggested that manipulation and mobilisation demonstrated similar effects on pain, function and patient satisfaction and they were found to be as effective as medication and acupuncture but better than a control and TENS at immediate and short-term follow up. Consequently, manual therapy is recommended for use in clinical practice (Child et al. 2008, Gross et al. 2010). A recent RCT compared the effectiveness of three manual therapy techniques: high velocity, low amplitude (HVLA), mobilisation and sustained natural apophyseal glide (SNAG) in patients with chronic neck pain (n=51) (Perez et al. 2014). At three months follow-up, all groups had a reduction of pain and disability. However, there was no statistically significant differences in the mean change between the groups. The authors concluded that there was no superiority of HVLA, mobilisation or SNAG in the short term (3 months) for patients with NS-NP.

3.3.1.2 Massage therapy

Massage is one of the oldest therapeutic remedies and uses the manipulation of the soft tissue as the main therapeutic tool (Sherman et al. 2014). Massage was investigated for pain and function for patients with NS-NP in four systematic reviews (Haraldsson et al. 2006, Vernon et al. 2007, Gross et al. 2007, Ezzo et al. 2007). These reviews found no strong or moderate evidence to support the effectiveness of massage for the management of patients with NS-NP. Major methodological issues such as lack of uniform definition of the massage technique, dosage and the mode of performance were often identified. Consequently, massage is not recommended for use as a single intervention when managing patients with NS-NP, it may be useful however if massage is utilised in combination with other active approaches such as exercise (Moffett and McLean 2006). One additional recent RCT (Sherman et al. 2014) investigated the ideal dose of massage on pain and function and concluded that 1 hour of massage 2-3 times a week was found to be the optimal dose for patients with mild NS-NP.

3.3.1.3 Electrotherapy

Electrotherapy is an umbrella term encompassing a number of physical modalities, such as Galvanic Current, Electrical Nerve Stimulation (ENS), Transcutaneous Electrical Nerve Stimulation (TENS), Pulsed Electromagnetic Fields Therapy (PEMF), and Repetitive Magnetic Stimulation (rMS). Electrotherapy methods were investigated for pain and function for patients with NS-NP in five systematic reviews (Kroeling et al. 2005, Gross et al. 2007, Hurwitz et al. 2008, Gross et al. 2010, Kroeling et al. 2013). Limited evidence supported the effect of rMS on pain and function in chronic NS-NP as a short-term treatment option, and inconsistent evidence supported the benefits of TENS for chronic NS-NP (Hurwitz et al. 2008). Meanwhile, limited evidence supported the benefits of PEMF at extremely low frequencies and high frequencies on pain reduction in patients with acute or chronic NS-NP immediately post treatment (Kroeling et al. 2005). Very low-quality evidence from a recent Cochrane review found that PEMF, rMS, and TENS are more effective than a placebo for pain reduction, while no benefit was found for all the other electrotherapeutic modalities (Kroeling et al. 2013). Overall, the effectiveness of electrotherapeutic treatment modalities remains uncertain when used as a single intervention for the management of NS-NP due to problems related to low methodological quality and funding bias (Gross et al. 2007; Hurwitz et al. 2008, Kroeling et al. 2013).

3.3.1.4 Acupuncture

Acupuncture is considered to be an alternative and complementary approach for the management of NS-NP, which requires the insertion of needles into the body for the purposes of pain reduction and/or to induce anaesthesia. Various approaches related to different countries including China, Japan and Korea have been proposed and tested (Trinh et al. 2006). Acupuncture was investigated for pain and function for patients with NS-NP in four systematic reviews (Trinh et al. 2006, Gross et al. 2007, Fu et al. 2009, Gross et al. 2010). Trigger point acupuncture was found to be more effective than other types of acupuncture on pain relief in the immediate post treatment phases and at short-term follow up (Itoh et al. 2007). Meanwhile strong to moderate evidence supported the effectiveness of acupuncture when compared to other inactive treatment approaches for pain relief immediately after treatment and at short-term follow up in patients with chronic NS-NP (Trinh et al. 2006, Vas et al. 2006, Gross et al. 2007). In addition, moderate evidence from a recent systematic review and one high quality RCT supported the effectiveness of acupuncture when compared to a control and a placebo on pain in the immediate post treatment phase and at short-term follow up (Fu et al. 2009, Liang et al. 2011). Overall, there is evidence to support the efficacy of using acupuncture to treat pain at short-term follow-up, but important clinical information such as the frequency and exact points to be acupunctured were not available. Consequently, it is not recommended as a single intervention for the management of patients with NS-NP (Moffett and McLean 2006, Child et al. 2008).

3.3.1.5 Traction

Traction is a passive treatment approach that involves a longitudinal stretch to the cervical spine. This stretch can be applied manually or mechanically for a specific period of time (continuous or static) or intermittently (on/off mode) (Graham et al. 2008). Cervical traction was investigated in three systematic reviews, which reported limited or no evidence to support the effectiveness of traction for NS-NP, and therefore any possible benefits of traction remain uncertain (Graham et al. 2006, Gross et al. 2007, Graham et al. 2008). Cervical traction is not recommended as a single intervention for the management of patients with NS-NP (Moffett and McLean 2006, Child et al. 2008).

3.3.2 Active treatment approaches

3.3.2.1 Therapeutic exercise

Therapeutic exercise is the execution of pre-arranged physical movement and activity related to the cervical spine region for the purposes of pain reduction, restoration of function and improved quality of life for patients suffering from NS-NP (Kay et al. 2005). It is an integral part of any rehabilitation programme and should be a key element for any physiotherapy approach when managing any musculoskeletal conditions including NS-NP (Child et al. 2008, Jull et al. 2008). Therapeutic exercise is frequently classified into general exercise for the neck and upper limb, cervical strengthening, cervical stretching, stabilising, endurance, balance and proprioception exercise. These components have been investigated as possible interventions for NS-NP in several systematic reviews (Sarig-Bahat 2003, Kay et al. 2005, Gross et al. 2007, Hurwitz et al. 2008). Moderate strength evidence supported the effectiveness of stretching and strengthening exercises for the neck and upper limb in the short and longterm on pain and function for chronic NS-NP (Kay et al. 2005, Gross et al. 2007, O'Leary et al. 2007, Ylinen et al. 2007, Hurwitz et al. 2008). Low to moderate strength evidence from a high-quality review suggested that strengthening exercises demonstrate optimal improvement in all outcomes (Kay et al. 2012). The addition of stretching and/or aerobic exercise should enhance a treatment programme for sub-acute and chronic NS-NP (Bertozzi et al. 2013). In their recent review and meta-analysis, O'Riordan et al. (2014) reported that active strengthening, stretching and aerobic exercise 3 times a week for approximately 30 to 60 minutes with 80% of maximum muscles voluntary contraction demonstrated the optimal benefits on pain, function and quality of life for patients with NS-NP (O'Riordan et al. 2014). Moderate strength evidence supported the short-term benefits of eye-fixation and neck proprioceptive exercise on pain and patient satisfaction (Sarig-Bahat 2003, Kay et al. 2005, Gross et al. 2007). Conflicting evidence were found regarding the effectiveness of home exercise (not supervised), group exercise and neck school (Kay et al. 2005, Gross et al. 2007, Hakkinen et al. 2008, Hurwitz et al. 2008).

3.3.2.2 The Mechanical Diagnostic and Therapy (MDT)

MDT, sometimes known as the McKenzie approach, is a comprehensive diagnostic and treatment approach commonly utilised for the management of back and neck pain. The approach is based on the individual's response to repeated movement or sustained posture in a specific direction (Clare et al. 2004). Self-management is the main

objective and the patient is educated with regards to the beneficial effects of repeated movement in a specific direction and the adverse effects of any movement in the opposite direction (Klaber Moffett et al., 2006). Despite the popularity of the McKenzie approach among UK physiotherapists (Foster et al. 1999), it was investigated in one systematic review (Clare et al. 2004). Research supporting the effectiveness of the McKenzie approach is limited (Moffett et al. 2006), which does not support or refute its effectiveness for patients with NS-NP.

3.3.2.3 Feldenkrais

Feldenkrais is an alternative and complementary treatment often used for managing chronic physical conditions including NS-NP. The theory of this approach is based on that patients with musculoskeletal conditions should have positive attitude towards function in order to reduce the potential for the development of chronic pain (Bearman and Shafarman 1999, Lundblad et al. 1999). The effectiveness of Feldenkrais remains unclear due to the limited research regarding its benefits for patients with NS-NP.

3.3.3 Patient education

Therapeutic patient education is a treatment approach, in which different techniques such as oral, written, or audio-visual techniques are used to provide patients with the necessary information and skills to manage their life with a disease (WHO 2001). Various education programmes, which were delivered orally or in a written/audio-visual form have been tested in a number of reviews and RCTs as an intervention for patients with NS-NP (Gross et al. 2007, Vernon et al. 2007, Hurwitz et al. 2008, Haines et al. 2008, Bernaards et al. 2008). Moderate evidence supported the benefits of education and/or counselling on pain and disability in the short- and medium-term follow up for female computer workers with NS-NP (Vernon et al. 2007). No evidence was found to support the benefits of traditional neck school when compared to no treatment for patients with NS-NP (Gross et al. 2007, Haines et al. 2008, Hurwitz et al. 2008). Overall, there is no evidence to support the effectiveness of patient education as a single intervention for the management of patients with NS-NP (Haines et al. 2008, Gross et al. 2012). However, clinical guidelines recommended that patient education should be used as part of a multimodal approach in the holistic management for patients with NSNP (Child et al. 2008).

3.3.4 Multimodal management approach

A multimodal approach to management is a combination of at least two interventions, with therapeutic exercise being a key component of any combination for the management of patients with NS-NP (Gross et al. 2007, Hurwitz et al. 2008). Overall, a multimodal management approach that combines stretching/strengthening exercise (supervised) and manual therapy (manipulation/mobilisation) with/without patient education is the optimal approach with strong evidence of effectiveness on pain reduction, improved function and patient satisfaction in the short- and long-term benefits for patients with NS-NP (Gross et al. 2002, Gross et al. 2004, Kay et al. 2005, Gross et al. 2007, Hurwitz et al. 2008, Kay et al. 2012). Moderate evidence suggested that a multimodal approach combining manual therapy and electrotherapy, medication or any other non-invasive approach demonstrated no difference on pain, function or patient satisfaction (Gross et al. 2004). A multimodal modal approach combining manual therapy with patient education or home exercise demonstrated no benefits on pain or function in patients with NS-NP (Gross et al. 2007, Hurwitz et al. 2008). Moderate evidence supported the effectiveness of a multimodal approach that includes exercise and manual therapy on pain and quality of life when compared with either exercise of manual therapy alone at the short-term follow up (Miller et al. 2010).

3.4 Conclusion

This review was conducted to establish current recommended non-invasive and nonsurgical treatment approaches for the management of patients with NS-NP as well as determining the effectiveness of these approaches when used as a single intervention or when combined in a multimodal management approach. Various active and passive treatment approaches are currently recommended for the management of patients with NS-NP, alongside the multimodal management approach which utilises at least two different approaches concurrently for the management of patients with NS-NP. However, strong evidence of effectiveness was only found for the multimodal approach which incorporate supervised stretching/strengthening neck and upper limb exercise and manual therapy (manipulation and mobilisation) with/without a patient education programme for acute/sub-acute or chronic NS-NP. The effectiveness of passive treatment approaches such as manual therapy, electrotherapy, massage, acupuncture or traction remains uncertain when used as a single intervention. In addition, it is not recommended to utilise these passive approaches in isolation, since they may lead to

patient passivity, inactivity, which consequently may contribute to disability behaviour (Swenson 2003, Moffett and McLean 2006).

The next chapter is a national survey which addresses the second research question of this thesis (see Section 1.5) and describe current UK physiotherapy practice in relation to the measurement and management of NS-NP. The findings of this current chapter informed the development of the survey instrument.

Chapter 4: Physiotherapy management of patients with non-specific neck pain: A national survey of current UK practice.

4.1 Introduction

This chapter describes a national survey investigating physiotherapy management and measurement of patients with non-specific neck pain (NS-NP).

Information about current physiotherapy practice in relation to the management and measurement of NS-NP is important for three reasons. First, to understand which management approaches are preferred by UK physiotherapists and the extent to which these approaches are supported by evidence of effectiveness. Second, to gain an insight into the current level of measurement in the assessment and during the management of patients with NS-NP (e.g. which OMs are most often used, which of the relevant constructs are frequently measured and what are the reasons for using or not using OMs when assessing or rehabilitating patients with NS-NP), and to determine the extent of the use of standardised OMs for NS-NP. For the purpose of this study OM is defined as an instrument, tool, task or questionnaire that is used in clinical practice or research to determine the presence of a condition/disease and measure its severity (Nelson and Berwick 1989, Duckworth 1999, Haigh et al. 2001). Moreover, this instrument should have the capability to objectively and/or subjectively detect and quantify changes in the construct of interest during and after rehabilitation (Abrams et al. 2006, Jette et al. 2009). Finally, the knowledge about current UK physiotherapy practice for NS-NP would highlight the impact of evidence on clinical practice and could be used to determine the extent of evidence-based dissemination and identify barriers to implementation. This could facilitate the development of new strategies to disseminate and/or implement evidence-based practice, if required, and enable the recommendation of priorities for future research.

Chapter 3 of this thesis summarised evidence of currently recommended treatment approaches for the management of NS-NP. Strong evidence of the benefits was found for therapeutic exercise of specific types when used in combination with other management approaches. A multimodal management approach is a combination of at least two interventions, with therapeutic exercise being a key component of any combination (Gross et al. 2007, Hurwitz et al. 2008). Multimodal management approaches are frequently categorised as: (1) exercise, manual therapy and patient education; (2) exercise and manual therapy; and (3) exercise and any other management

approach. Despite the commonality and burden of NS-NP, and the availability of a wide range of management approaches, there is limited information regarding the patterns of incorporating these approaches into UK physiotherapy practice.

Upper limb disability (ULD) is frequently associated with NS-NP (Daffner et al. 2003, Falla et al. 2004, Frank et al. 2005, Bot et al. 2005). Optimal rehabilitation of NS-NP requires evaluation of the upper limb using suitable OMs and should include ULD rehabilitation in the management plan if indicated (McLean et al. 2011, Osborn and Jull 2013). Chapter two of this thesis investigated all available OMs that have been used to measure UL functional capacity in patients with NP. The findings reported that there are four promising instruments, however significant methodological and quality issues prevented a clear recommendation for any of these OMs (Alreni et al. 2017). Limited evidence suggested that UK physiotherapists may be inadequately evaluating their patients with NS-NP (McLean et al. 2011). The extent to which validated NP, ULD, or other potentially relevant OMs are used by UK physiotherapists when managing NS-NP is unknown.

4.1.1 Aim

The overall aim of this national survey was to investigate current UK physiotherapy practice in relation to the measurement and management of NS-NP. The objectives were to:

1. Describe current UK physiotherapists with regard to their management of patients with NS-NP, utilisation patterns of the multimodal management approaches and rehabilitation of the ULD.
2. Describe current practice of UK physiotherapists with regard to their utilisation of OMs. Specifically of interest within this thesis is utilisation of ULD OMs in the assessment and during the rehabilitation of patients with NS-NP.
3. Exploring the relationship between demographic characteristics of participants and management strategies utilised as well as OMs utilised.

4.2 Methods

4.2.1 Design

This online web-based survey explored UK-based physiotherapists' use of management approaches and OMs utilisation for patients with NS-NP. The survey instrument was designed and extensively evaluated to ensure robust face and content validity as well as

acceptability and feasibility. An extensive online methodology utilising Social Networking Sites including Twitter, Facebook and LinkedIn was used in the survey to optimise recruitment of participants. The survey was conducted from March 2016 to November 2016 and approved by Sheffield Hallam University, Faculty Research Ethics Committee (see Appendix 4).

4.2.2 Rational for online web-based survey

In order to meet the aim of this investigation, an online web-based survey was used to develop a coherent picture of current UK physiotherapeutic management and measurement of NS-NP. This research method was used in this national survey for four main reasons. First, it provides the necessary cost-effective method of data collection, in which a large sample size can be achieved at very low cost compared to other survey methods, such as a postal survey or a telephone survey (Couper 2000, Dillman 2000, Shannon et al. 2001). This large sample size can lower the sample variance and provide the potential for sub-group analysis, which can strengthen the power of the study and thus the validity of the findings (Cook et al. 2000, Couper et al. 2004, Manfreda et al. 2008). Second, it saves time and provides the required speed as well as coverage over a wide geographical area for data collection, since it enables instant distribution and continuous data collection (i.e. it can reach respondents wherever they are and at whatever time is most convenient to them) (Couper et al. 2001). Third, it provides the needed speed and accuracy of data analysis, since it allows instant access to participants' responses that can be automatically downloaded to an electronic database or statistical package, eliminating human error in data coding and data entry (Crawford et al. 2005, Dillman 2007). Finally, it enables innovative questionnaire design to be developed using advanced and interactive features (e.g. screen design, text, question presentation, respondents' response format and the survey navigation). In addition, it allows the use of drop-down boxes, pop-up windows, routing systems (i.e. the system enables respondents to skip portion of the survey which is not relevant to their practice) and progress indicators in the design, features which are not always available with the other survey modes.

These features provide a dynamic survey process that ensures the acceptability, feasibility, simplicity and brevity of the survey instrument and have been shown to increase survey response rates (Dillman et al. 1998, Fricker and Schonlau 2002, Dillman and Smyth 2007, Manfreda et al. 2008).

4.2.3 Disadvantages of web-based survey

4.2.3.1 Access to the internet

Missing out respondents who do not have internet access is the first challenge facing this approach. However, in the last decade, the internet has become a key component in the navigation of everyday life (Hughes et al. 2012). Meanwhile, the data produced by the UK Office for National Statistics annual report for the year 2016 estimated that 87.9% (45.9 million) of the population had recently (in the last 3 months) used the internet; almost all adults aged 16-24 years were recent internet users (99.2%); 89.4% of men (22.8 million) and 86.4% of women (23.1 million) were recent internet users (Office for National Statistics 2016). Furthermore, evidence from the literature suggested that the majority of healthcare professionals in the UK including physiotherapists are using popular Social Networking Sites such as the Facebook, Twitter, LinkedIn and YouTube for social interaction and information exchange as well as business advertising (Ahmed et al. 2012). This indicates that the majority of UKbased physiotherapists are likely to be internet users and thus overcomes the first main reported challenge regarding internet access.

4.2.3.2 Low response rate

Web-based surveys tend to have poorer response rate, which is reported to be approximately 11% lower than that of other survey modes such as postal surveys or telephone surveys (Couper 2000, Fan and Yan 2010). Poor response rate is a major challenge threatening the validity of the survey's findings (Couper et al. 2004). In order to overcome this challenge, and to maximise the number and diversity of respondents the following steps were taken. Firstly, the survey instrument was developed and designed in accordance with the Web-Based Survey Design Standards reported by Crawford et al. (2005), which provide guidance for developing a robust web-based survey design that ensure acceptability and feasibility of the survey instrument. This includes standards to be used when designing the visual display of questions, responses and the supporting survey materials to the respondents; question presentation; respondents' input/response formats; and the survey navigation/interaction. In addition, an extensive online methodology incorporating popular Social Networking Sites including Twitter, Facebook and LinkedIn was developed and implemented in the survey procedure (see section 4.2.6).

4.2.4 Sampling

4.2.4.1 Sampling method

The main UK physiotherapy professional bodies, the Health and Care Professional Council (HCPC) and the Chartered Society of Physiotherapy (CSP) do not allow access to their registrants for the purpose of recruiting participants for research projects or surveys. Meanwhile, accessing physiotherapists working in the NHS which comprised approximately 60% of all UK-physiotherapists requires approval from each selected National Health Services (NHS) Trust prior to the survey administration, which is time consuming and is also not representative of the total population of physiotherapists. Furthermore, other UK physiotherapy organisations such as the Musculoskeletal Association of Chartered Physiotherapists (MACP) and the Physiotherapy Pain Association (PPA) are also not representative of the total population and therefore not considered feasible for a national survey. Hence, a non-probability sampling approach was used in this PhD national survey. In to reduce the likelihood of sampling bias and to recruit a large sample, the survey procedure utilised an extensively pre-designed innovative online methodology incorporating popular social networking sites including Twitter, Facebook and LinkedIn. The rationale and details of the online methodology used in Twitter, where the majority of responses were achieved is presented in section 4.2.6.

4.2.4.2 Sampling frame

There are 48,611 physiotherapists registered in the UK (HCPC, 2016). More than 98% of these registrants are also members of the Chartered Society of Physiotherapists (CSP), (the professional, educational, and trade union representing UK physiotherapists). Approximately 60% of these chartered physiotherapists work for the publicly funded NHS in the UK. The remaining 40% work in a variety of settings, such as private clinics, private hospitals, military hospitals, sports clubs, or teaching in higher education institutes. UK-based musculoskeletal physiotherapists who are involved in the management of NS-NP were the sample frame (target population) in this national survey, though the size of this sampling frame is unclear.

4.2.4.3 Sampling procedure

Physiotherapists were eligible to participate in this survey if they were: currently a member of the HCPC; working in either an NHS setting, non-NHS setting (e.g. private clinics, private hospitals, military hospitals or sports clubs), or a combination of NHS

and non-NHS settings; and have seen at least one case of neck pain (NP) in the last 6 months. Physiotherapists working in areas such as intensive care, mental health, respiratory care, paediatrics, stroke services, elderly care or inpatient settings; currently not involved in the management of patients with NP; or not practicing in the UK were ineligible to participate. From the sampling approach used in this survey, there was no way of identifying those UK musculoskeletal physiotherapists who are involved in the management of NP. Therefore, eligibility conditions were indicated in the survey's invitation and the eligibility decision was left to the discretion of participants. Written consent was not sought from each participant; however, consent was assumed if physiotherapists completed the survey (Crawford et al. 2005).

4.2.4.4 Sample size

The sampling method (non-probability) used in this survey meant that the number of potential participants could not be reliably determined, and consequently it was not possible to estimate the sample size and/or response rate. However, an innovative online methodology was used in the survey procedure to maximise the number as well as the diversity of participants in this survey.

4.2.5 Survey instrument

For the purpose of this study a new survey instrument was developed and designed to capture information regarding current UK physiotherapy measurement and management of NS-NP. The survey instrument (see Appendix 5) included a participant information sheet and thereafter comprised questions divided into three sections: (1) demographic characteristics of physiotherapists, (2) treatment approaches that are used when managing NS-NP, and (3) OMs that are used in the assessment and/or management of NS-NP. The rationale and details about each section are presented in sections 4.2.5.3 to 4.2.5.5. The development and validation of the survey instrument were iterative with multiple revisions and piloting procedures to ensure face and content validity (Dillman et al. 1998, Couper 2000, Couper et al. 2001, Dillman 2007, MacDermid et al. 2013, Carlesso et al. 2014).

Two reviews of the literature were undertaken to inform the preliminary development of the survey instrument. First, a literature review of all available treatment approaches for the management of NS-NP was conducted to determine what conservative, non-invasive, management approaches have been investigated and currently recommended for clinical practice (see Chapter 3). Available management approaches were

thematically grouped to produce an appropriate number of items and response options that represent the range of interventions currently used in clinical practice for patents with NS-NP. Details are presented in section 4.2.5.4. Second, a systematic review of all available upper limb OMs for patients with NP (see Chapter 2) was conducted to identify, critically examine and recommend a list of suitable upper limb OMs developed or validated for patients with NP. Furthermore, evidence from additional systematic reviews, development and/or validation studies of all available OMs for patients with NP were sought (Pietrobon et al. 2002, de Koning et al. 2008, Dvir and Prushansky 2008, Siva et al. 2010, Terwee et al. 2011, Schellingerhout et al. 2012, Horn et al. 2012). All the identified OMs of ULD for patients with NP alongside all available NP OMs were collated and thematically grouped to represent the spectrum of OMs for NP and its associated ULD. Details are presented in section 4.2.5.5.

4.2.5.1 Development

The survey instrument was developed in accordance with the Web-Based Survey Design Standards reported by Crawford et al. (2005), which published standards developed from theory and practice with regard to the screen design, text, question presentation, respondents' input/response formats, and survey navigation/interaction to ensure acceptability and feasibility as well as simplicity and brevity of a survey, which has been shown to increase the response rate of web-based surveys (Baruch 1999, Baker et al. 2003, Couper et al. 2004, Couper et al. 2007, Fan and Yan 2010). Figure 4.1 shows the procedure that was used in the development and validation of the survey instrument.

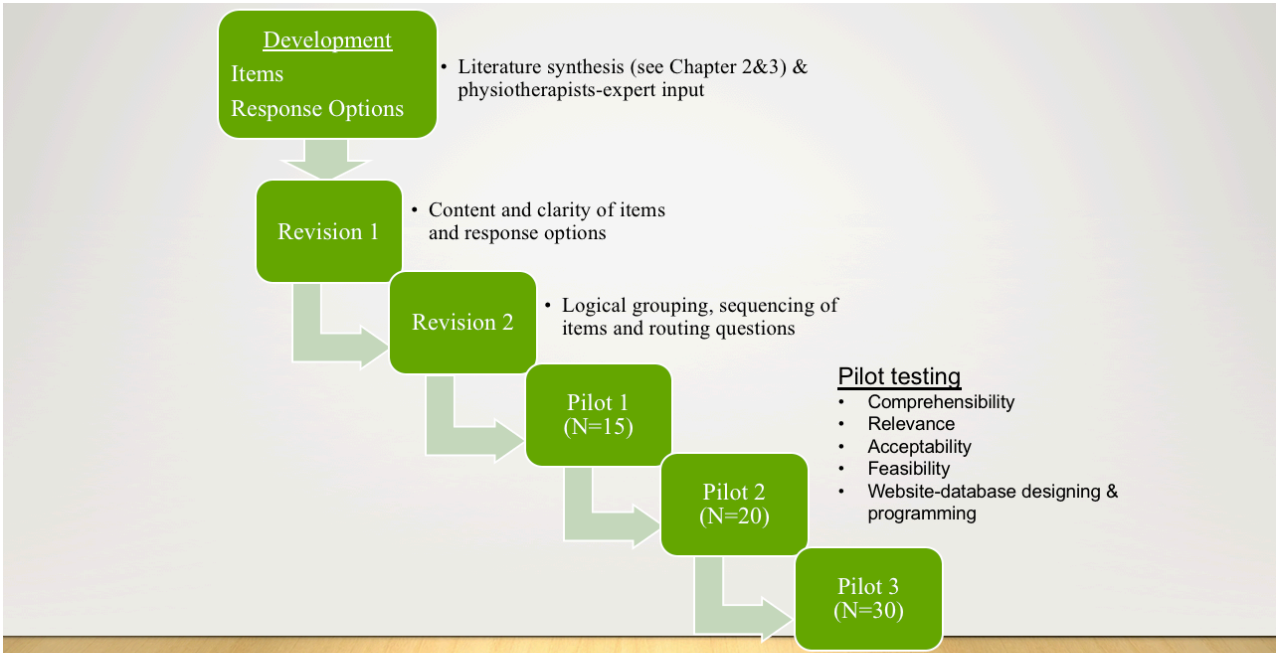


Figure 4.1 Survey Instrument (Development Process)

4.2.5.2 Validation

The survey instrument underwent two rounds of revision and three rounds of pilot testing. The initial revision stage emphasised the content and clarity of each individual item and response option and resulted in the inclusion of additional definitions and examples for each OM in section 3. In addition, the wording was modified for 2 questions and 5 response options in sections 2 and 3. The final revision stage emphasised the logical grouping, sequencing of items and routing (filter) questions that enabled respondents to skip portions of the survey which were not relevant to their practice. This stage resulted in the re-ordering of several items and response options in sections 2 and 3.

The initial pilot testing was conducted in three stages: stage 1 (n=15) and stage 2 (n=20), involved experts/clinical physiotherapists from the target audience. Participants were asked to review the survey instrument with regard to the electronic format, the routing questions and the functionality of the web-page design and programming. Finally, field testing was conducted involving physiotherapists, experts, working in clinical practice and research practice, alongside clinical physiotherapists from the target population (n=30). These experts and clinical physiotherapists were requested to review the survey instrument for accuracy, clarity, completeness and burden. Field testing resulted in several modifications to clarify the aim of each section, improve the wording of sentences and improve the organisation of the survey content.

Participants in this survey were able to select all the management approaches (section 2) and the OMs (section 3) that they would use most often when managing their patients with NS-NP. This was to enable the identification of UK physiotherapists who were using multimodal management and/or measurement approaches. Progress indicators were used at the bottom of each screen to inform respondents of their progress throughout the questionnaire and to prevent drop out before completing the survey (Jeavons 1998, Couper et al. 2001). Closed questions only were used in this survey to reduce the time burden (Crawford et al. 2005). However, respondents were able to identify any 'other' management approaches or OMs that they would use but which were not identified in the response option lists.

4.2.5.3 Section 1 demographic characteristics

Questions about physiotherapists' demographic characteristics were included in the survey instrument to permit a full description of the sample, examine the diversity of the sample, compare the sample to the total population and finally to enable examination of possible associations between demographic characteristics and the selection of management approaches and/or utilisation of OMs. The demographic characteristics of physiotherapists (e.g. gender, setting, years of experience, nation where physiotherapists were clinically practicing, postgraduate training level, caseload information and special interest in NS-NP) were obtained using standard closed format questions. Drop down boxes allowing one answer only were used to provide the response options.

4.2.5.4 Section 2 management approaches

Questions about management approaches were included in the survey to enable full description of current UK physiotherapy practice in the management of NS-NP, which is the first objective of this national survey (see Section 4.1.1). Selection of the items relating to upper limb management strategies were used to evaluate/establish the level of upper limb rehabilitation undertaken during the management of NS-NP. Section 2 was launched with a statement clarifying the aim of the section "In this section, we are interested in identifying the management approach/approaches you typically use for patients with NS-NP". This was followed by a clear definition of NS-NP, "NS-NP is defined here as a dysfunction in the cervical structure NOT caused by any serious acute trauma (e.g. Whiplash Associated Disorder), systemic disease, neurological disorder (e.g. Cervical Radiculopathy, Nerve Root Compression) or inflammatory condition". This section enabled respondents to select as many management approaches as they would most often use to identify those respondents who use multimodal management approaches. Selecting management approaches such as therapeutic exercise, manual therapy and/or electrotherapy triggered another question and response options, via a drop-down box, requesting those respondents to select the component/method they typically used most often "Which component/components – method/methods do you use regularly for patients with NS-NP?". Response options in this section were presented as follows: therapeutic exercise followed by its main components (e.g. general aerobic/ strengthening/endurance exercise, cervical strengthening exercise, upper limb strengthening exercise, cervical stretching exercise, upper limb stretching exercise, cervical stabilising exercise, upper limb stabilising exercise, balance exercise, proprioception exercise for the eyes, proprioception exercise for the cervical spine and

proprioception exercise for the upper limb); manual therapy followed by its main methods (e.g. Maitland, Mulligan, Society of Orthopaedic Medicine and Manipulation (grade V); and electrotherapy modalities followed by its main methods (e.g. Galvanic Current (DC), Electrical Nerve Stimulation (ENS), Pulsed Electromagnetic Fields (PEMF), Transcutaneous Nerve Stimulation (TENS) and Repetitive Magnetic Stimulation (rMS)). These management approaches were followed by other interventions such as the McKenzie method; therapeutic patient education; massage therapy (all types); acupuncture; traction; heat/cold; taping/strapping; hydrotherapy; Feldenkrais; and other management approach/approaches.

4.2.5.5 Section 3 outcome measures (OMs)

Questions about OMs were included in this survey to permit a full description of current UK physiotherapy practice in the measurement of NS-NP, which is the second objective of this national survey (see Section 4.1.1). Selection of items relating to physical/functional upper limb OMs were used to establish the level of upper limb evaluation undertaken in the assessment and during the management of NS-NP. This section was launched with a routing question to reduce the time burden and prompt brevity of the survey instrument "do you use OMs in the assessment/management of patients with NS-NP?". If respondents selected NO, they were asked, using a dropdown box, to indicate their reasons for not using OMs and then routed to the end of the survey where they were thanked for participating in the survey. Selecting YES triggered another 3 questions and response options, using a drop-down box, requesting those respondents to select the reasons for using OMs, the frequency with which they use any measures (i.e. routinely = >70% of cases, regularly = 51-70% of cases, sometimes = 1150% of cases and rarely = <10% of cases) and the OMs that they use most often as well as the patterns of use. Response options in this final section were structured and presented as follows: routinely = >70% of cases, regularly = 11-70% of cases, rarely = 1-10% of cases and never = 0% of cases (Crawford et al. 2005).

Items in this section were presented as follows: first, Patient-Reported OMs (PROMs) (i.e. pain measures, physical functioning OMs, work status OMs, psychological distress OMs and quality of life OMs [global OMs and generic multidimensional OMs]). Second, Performance-Based OMs (PBOMs) (i.e. pain threshold perception, motion OMs, muscle function OMs and functional performance OMs).

4.2.6 Survey procedure

A domain name "UK-neckpainsurve.com" and hosting web were purchased. A webpage and associated database were designed and programmed for data collection in this web-based survey. An extensive and innovative online methodology incorporating popular social networking sites including Twitter, Facebook and LinkedIn was developed and implemented for the recruitment and administration of this national survey between August 2016 and November 2016. This was to increase the number, as well as the diversity, of participants. Details of the strategy used in Twitter, where the majority of responses were achieved, are presented in the paragraph below.

Strategy development was an iterative process with up-front loaded online materials. First, a credible and professional public Twitter profile as a PhD student at Sheffield Hallam University was created using a friendly bio name (@ResearchingNeck) and a JPEG photo of the researcher. A public Twitter profile was used to enable all Twitter users to view and interact with the tweets. Second, a banner containing Sheffield Hallam University's logo, JPEG photo of the director of study (Dr Sionnadh McLean) and another JPEG photo of the researcher alongside a concise, friendly and professional survey invitation were designed (see Appendix 16). Third, influential UK-based individuals, groups and organisations associated with Musculoskeletal Physiotherapy, as well as a high number of UK-based Musculoskeletal Physiotherapists on Twitter were identified using the Social Media Analytics Software (Followerwonk – A Moz App) and then followed by the researcher. Fourth, several Twitter posts using wording suitable for each identified individual, group or organisation were created using other public Twitter accounts after gaining permission from the account holders. Fifth, key posting times (that is, the times of the day when UK-based physiotherapists have just signed-in Twitter) were identified using the aforementioned analytics software. Finally, tweets were posted to these identified and followed individuals, groups and organisations followed on a daily basis for 3 months.

In Twitter, a tweet could contain a maximum of 140-characters (in the time when the survey was conducted) but JPEG and GIF photos did not count towards the character limit. Thus, a banner was included, which contained a GIF photo and JPEG photos of the researcher and the director of study alongside the survey invitation in every tweet. All tweets began with an @username of one of the identified influential individuals, groups or organisations. This was to enable their followers to immediately see the tweet

in their home timeline. Individuals, groups and organisations were also asked to retweet or quote tweet the tweet (that is, retweet with added comment requesting followers to complete the survey). Since Twitter users often miss tweet posted when they are not online, particular times of the day were targeted; for example, when physiotherapists have just have signed-in (that is, 7:10am, 10:40am, 17:45 and 21:50). A tweet begins with @username, a friendly request to complete and retweet the survey, the survey obtained google link http://goo.gl/OynKlq and the survey's banner was posted to every identified individual, group and organisation. Since groups and organisations were being asked to retweet and/or quote the tweet about this study, the process repeated was repeated every two weeks. The researcher performed "like" to every retweet or quote tweet in the following day after retweeting at one of the key posting times. The researcher also posted a thank you tweet to those who retweet or quote tweeted the original tweet in the second day after retweeting or quote tweeting the original tweet at one of the key posting times.

In this strategy, posting tweets to the identified influential individuals, groups and organisations was considered to constitute recruitment of participants and administration of the survey instrument since those tens of thousands of physiotherapists who follow these individuals, groups and organisations should have seen/received the tweet associated with the study, which included all the materials about the survey, in their home timeline immediately after posting. "like" by the researcher to the tweet or quote tweet in the following day was considered to be the first reminder since it enabled the survey to be seen/received again in the home timeline of those followers. "thanks" tweet, which was posted by the researcher in the second day was considered to be the second reminder since it enabled the survey to be seen/received again, for the third time, in the home timeline of the followers. "like" to our "thanks" tweet by the individual, group and organisation, which always happened, was considered to be the third reminder since it enabled the survey to be seen/received again for the fourth time, in the home timeline of the followers. Overall, this recruitment strategy supported access to a national population of UK physiotherapists with a wide range of demographic characteristics.

4.2.7 Ethical consideration

In survey research methods, two main ethical issues always arise with regard to anonymity and consent. In this PhD national survey, registration to complete the survey

was not required, information which enables identification of respondents such as name, email address, telephone number and/or place of work were not requested at any stage of the data collection, and written consent from each respondent was not sought in order to ensure anonymity of the survey. The eligibility decision was left to the discretion of participants and consent was assumed if physiotherapists completed the survey (Couper 2000).

4.2.8 Data analysis

At the end of the survey period, data were collated and transferred from the web-based database into Microsoft Excel 2016 where data were checked. Subsequently, the data were transferred into SPSS (IBM SPSS Statistical Software, version 23.0) for statistical analysis.

In order to meet the objectives of this survey, descriptive analysis (frequencies, and percentages) was used to present the demographic information, the utilisation patterns of the management approaches and the utilisation patterns of OMs in the assessment and during the management of NS-NP. Subsequently, the utilisation patterns of multimodal management and measurement approaches, the reasons for using or not using OMs, and the patterns of evaluating and rehabilitating ULD were descriptively analysed.

Chi-square test, Phi and Cramer's V tests were used to examine the strength and significance of any association/differences in clinical practice between groups of UK physiotherapists (e.g. physiotherapists with different gender, years of experience, setting, postgraduate training level or practicing clinically in different nation), the utilisation patterns of multimodal management approach packages and OMs (Altman, 1991).

4.3 Results

4.3.1 Demographic characteristics

In total, 2101 physiotherapists who were members of the HCPC, practicing in the UK and involved in the management of patients with NS-NP completed the survey. The demographic characteristics of those physiotherapists are summarised in Table 4.1. Respondents were predominately female (57%), of which the majority (67%) were

working within the NHS (either exclusively in the NHS, or in a combination of NHS and non-NHS settings). A substantial component (44%) of respondents had 6-10 years of experience and the largest subgroup of physiotherapists practiced in England (66%). A slim majority were without postgraduate training and of the 48.7% of physiotherapists who reported that they had completed postgraduate training, 72.5% had completed an MSc, whereas 8.5% had completed a PhD. More than 40% of respondents had a caseload which included 25-50% patients with NS-NP and 65.1% of the sample reported that they had no special interest in NS-NP. In the item relating to "years of practice", because of small numbers, the categories "less than 2 years" and "2-5 years" were combined to create the category "0-5 years". Similarly, in the item relating to "proportion of patients with NS-NP", the categories "51-75%" and "75+%" were combined to create category "≥51%".

Table 4.1: Demographic characteristics of physiotherapists

Variables	N	%
Gender		
Female	1193	56.8
Male	908	43.2
Setting		
Exclusively in the NHS	1005	47.8
Exclusively in Non-NHS	628	29.9
Combination of NHS & Non-NHS	464	22.1
Other settings	4	0.2
Years of Practice		
0-5 Years	313	14.9
6-10 Years	928	44.2
11-15 Years	395	18.8
15+ Years	465	22.1
Nation		
England	1398	66.5
Scotland	297	14.1
Wales	285	13.6
Northern Ireland	121	5.8
Postgraduate Training		
With	1024	48.7
Without	1077	51.3
Proportion of Patients with NP		
<25%	881	41.9
25-50%	925	44.1
≥51%	295	14
Special Interest in NP		
No	1367	65.1
Yes	734	34.9

Notes: NHS: National Health Services, NP: Neck Pain

4.3.1.1 Demographic characteristics distribution

Demographic and participant characteristics (e.g. gender, setting and the nation, where physiotherapists practicing clinically) of the survey sample alongside the total UK physiotherapy population (HCPC, 2016) are presented in Table 4.2. Comparison between the survey sample and the UK physiotherapy population regarding the above demographics indicated a slightly higher proportion of the sample worked at least partly in the NHS (69.9%) compared with what is reported by the HCPC (60%). The sample had a lower proportion of female physiotherapists (56.8%) compared with those reported by the HCPC (78.2%). The sample was more balanced across the four UK nations with fewer located in England (66.5%) compared with the HCPC figures (83.3%). Since there was no demographic data available from the HCPC and/or the CSP regarding UK physiotherapists who are working in combination between NHS and non-NHS settings; in this survey, physiotherapists who reported working partly in the NHS were combined with those who reported working exclusively in the NHS. However, physiotherapists who reported working in other settings were combined with those who reported working exclusively in non-NHS during the above comparison. Overall, whilst the sample does not perfectly reflect the HCPC figures, a clear consistency is evident with the majority working in the NHS, being female and practicing clinically in England.

Table 4.2: Demographics of the survey population vs the UK physiotherapy population

Variables		UK (physiotherapy population)	Survey sample
Total number of physiotherapists		N=48611	N=2101
Setting	NHS	60.0%	69.9
	Non-NHS	40.0%	30.1
Gender: N (%)	Female	38012 (78.2)	1193 (56.8)
	Male	10596 (21.8)	908 (43.2)
Gender distribution in each nation: N (%)			
England	Total population	40455 (83.3)	1398 (66.5)
	Female	31375 (77.5)	712 (50.9)
	Male	9080 (22.5)	686 (49.1)
Scotland: N (%)	Total population	4198 (8.5)	297 (14.1)
	Female	3515 (83.7)	243 (81.8)
	Male	683 (16.3)	54 (18.2)
Wales: N (%)	Total population	2189 (4.5)	285 (13.6)
	Female	1706 (77.9)	213 (74.7)
	Male	483 (22.1)	72 (25.3)
Northern Ireland (NI): N (%)			
	Total population	1766 (3.6)	121 (5.8)
	Female	1416 (80.2)	25 (20.7)
	Male	350 (19.8)	96 (79.3)

4.3.2 Utilisation of management approaches

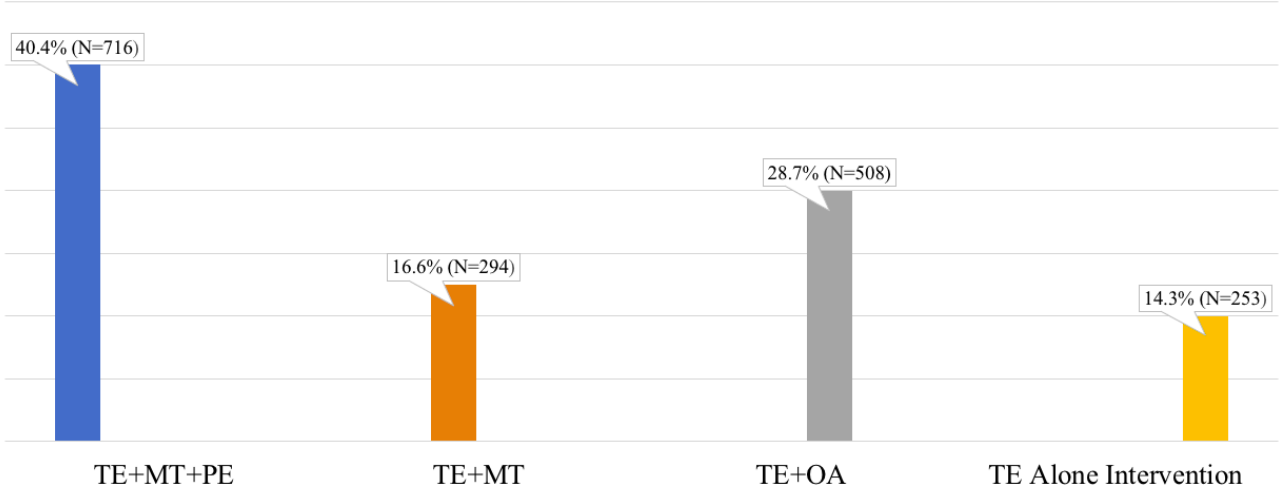
The management approaches that UK physiotherapists reported that they would use most often when managing their patients with NS-NP are summarised in Table 4.3. Therapeutic exercise of various kind was the most frequently used management approach, with 84.3% (n=1771) of physiotherapists reporting that they would use it most often when managing this patient group. However, 40.4% (n=716) of physiotherapists who were using therapeutic exercise reported that they would use it in combination with manual therapy and patient education programmes (multimodal approach package 1), 16.6% (n=294) of physiotherapists reported that they would use it in combination with manual therapy only (multimodal approach package 2), 28.7%% (n=508) of physiotherapists reported that they would use it in combination with any other management approach (multimodal approach package 3) and the remaining 14.3% (n=253) of physiotherapists reported that they would use therapeutic exercise as a standalone management approach when rehabilitating their patients with NS-NP (see Figure 4.2). The majority (60.7%) of respondents reported that they would use patient education programmes most often when managing patients with NS-NP. However, all physiotherapists who were using patient education reported that they would use it in combination with therapeutic exercise and manual therapy. Manual therapy methods were the most prevalent passive (hands-on) management approach, with 58.7% (n=1233) reporting that they would use it most often. However, 82% (n=1010) of physiotherapists who were using manual therapy reported that they would use it in combination with therapeutic exercise with/without patient education programmes, whereas 18% (n=223) of physiotherapists who were using manual therapy reported that they would use it as standalone management for their patients with NS-NP. Other active and passive management approaches (e.g. the McKenzie approach, massage therapy, acupuncture, heat/cold, taping/strapping and traction) were less commonly used. The electrotherapy methods, hydrotherapy and Feldenkrais, were not used by physiotherapists in this survey.

Table 4.3: Management approaches (N=2101)

Approaches	N	Percentage of Cases
Therapeutic Exercise	1771	84.3%
Therapeutic Patient Education	1275	60.7%
Manual Therapy	1233	58.7%
Massage Therapy	578	27.5%
Acupuncture	411	19.6%
Heat/cold	343	16.3%
Taping/Strapping	265	12.6%
The McKenzie Method	215	10.2%
Traction	158	7.5%
Other Management Approach	41	2.0%
Electrotherapy	0	0%
Hydrotherapy	0	0%
Feldenkrais	0	0%

Other Management Approaches Reported: Advice on Remaining Active = 4, Bobath = 2, Breathing Awareness practice = 1, Cognitive Functional Therapy (CFT) = 12, Depending on Presentation and Patient responses to Previous Treatment = 1, Education = 1, Intramuscular Trapezius Injections if is Warranted = 1, Psychological Intervention = 5, Relaxation Techniques = 6, Thoracic Spine Manipulation/Mobilisation = 1.

Figure 4.2 Utilisation of Therapeutic Exercise in Multimodal Approach Packages



TE+MT+PE = Therapeutic Exercise + Manual Therapy + Patient Education,
TE+MT = Therapeutic Exercise + Manual Therapy,
TE+OA = Therapeutic Exercise + Other Management Approach

4.3.2.1 Therapeutic exercise

The types of therapeutic exercise that physiotherapists reported that they would use most often in the management of patients with NS-NP are summarised in Table 4.4. In this survey, 15.7% (n=330) of physiotherapists reported that they would not use any type of therapeutic exercise when managing patients with NS-NP. The majority of the 84.3% (n=1771) of physiotherapists who were using therapeutic exercise reported that they would use cervical strengthening, stretching and general aerobic/strengthening/endurance exercise most often in the management of NS-NP. A

substantial proportion, 40.5% (n=717), of physiotherapists reported that they would use upper limb strengthening exercise when managing patients with NS-NP. Therapeutic exercise types such as cervical stabilising, proprioception for the cervical spine, upper limb stabilising, upper limb stretching and balance exercise were less commonly used. No "other" exercise components/types were identified.

Table 4.4: Breakdown of therapeutic exercise components used 84.3% (N=1771)

Components	N	Percentage of Cases
Cervical strengthening exercise	1348	76.1%
General aerobic/strengthening/endurance exercise	1222	69.0%
Cervical stretching exercise	1031	58.2%
Upper limb strengthening exercise	717	40.5%
Cervical stabilising exercise	629	35.5%
Proprioception exercise for the cervical spine	376	21.2%
Upper limb stabilising exercise	339	19.1%
Upper limb stretching exercise	289	16.3%
Proprioception exercise for the eyes	243	13.7%
Balance exercise	241	13.6
Proprioception exercise for the upper limb	146	8.2%
Other Components	0	0%

4.3.2.2 Manual Therapy

The manual therapy methods that UK physiotherapists reported that they would use most often when managing patients with NS-NP are summarised in Table 4.5. In this survey, 41.3% (n=868) of physiotherapists reported that they would not use any manual therapy method in the management of NS-NP. The Majority (73.3%) of physiotherapists who were using manual therapy reported that they would use the Maitland approach, whilst 45.5% reported that they would use the Mulligan approach most often when managing their patients with NS-NP. Other manual therapy approaches (e.g. Grade V manipulation, Kaltenborn, Cyriax and and the Society of Orthopaedic Medicine) were less commonly used. No "other" manual therapy methods were identified.

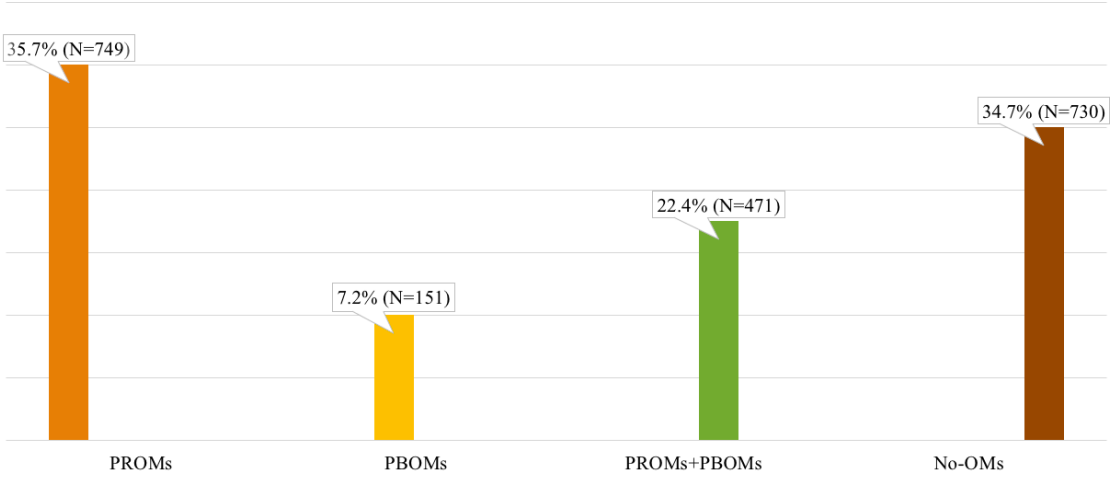
Table 4.5: Breakdown of manual therapy methods used 58.7% (N=1233)

Methods	N	Percentage of Cases
Maitland	891	72.3%
Mulligan	561	45.5%
Manipulation (Grade V)	124	10.1%
Kaltenborn	74	6.0%
Cyriax	54	4.4%
Society of Orthopaedic Medicine	51	4.1%
Other Methods	0	0%

4.3.3 Utilisation of OMs

Physiotherapists' current practice regarding their utilisation of OMs in the assessment and/or during the management of NS-NP is shown in Figure 4.3. One-third (34.7%, n=730) of physiotherapists in this survey reported that they would not use any OMs in the assessment/management of patients with NS-NP. However, 22.4% (n=471) of physiotherapists reported that they would use a multimodal measurement approach, which is a combination of Patient-Reported OMs (PROMs) and Performance-Based OMs (PBOMs). Meanwhile, 35.7% (n=749) of physiotherapists reported that they would use PROMs only, whereas only 7.2% (n=151) of physiotherapists reported that they would use PBOMs alone in the assessment/management of patients with NS-NP.

Figure 4.3: Utilisation of Multimodal Measurement Approach (N=2101)



PROMs: Patient-Reported Outcome Measures
PBOMs: Performance-Based Outcome Measures
OMs : Outcome Measures

4.3.3.1 Utilisation of PROMs

The PROMs that UK physiotherapists reported that they would use most often in the assessment/management of their patients with NS-NP are summarised in Table 4.6. Selection of ULD PROMs was used to determine the utilisation of ULD PROMs in the assessment and/or during the management of NS-NP. All of the PROMs were rarely used, although the most commonly utilised OMs were single dimensional numeric pain rating scales (i.e. the Visual Analogue Scale (VAS) and the Numeric Rating Scale (NRS). Quality of life OMs (i.e. Euro-Qol/EQ5D, WHOQOL-Brief and SF-36/SF-12), physical function OMs (e.g. NDI and PSFS), psychological distress OMs (e.g. fear of movement scale and depression/anxiety scale) were rarely, if ever, used. ULD PROMs (e.g. Disability of Arm, Shoulder and Hand (DASH/QuickDASH) and Neck and Upper Limb Index (NULI)) were almost always never used.

Table 4.6: Utilisation of PROMs 59.3% (N=1246)

Measures	Utilisation			
	Routinely >70%	**Regularly 11-70%**	**Rarely 1-10%**	**Never 0%**
Visual Analogue Scale (VAS)	36.9%	1.2%	3.4%	58.4%
Numeric Rating Scale (NRS)	18.4%	4.1%	0%	77.5%
Euro-Qol/EQ5D	14.8%	3.4%	1.3%	80.5%
Patient Specific Functional Scale (PSFS)	9.2%	4.7%	3.8%	82.3%
Neck Disability Index (NDI)	5.9%	3.2%	5.9%	85%
Fear of Movement Scales	7.9%	3.3%	2.8%	86.0%
Time Lost from Work	3.5%	4.5%	1.6%	90.4%
Depression/Anxiety Scale	3.6%	1.1%	4.6%	90.7%
Disability of Arm, Shoulder and Hand (DASH)/QuickDASH)	2.3%	1.0%	5.0%	91.6%
Pain Catastrophizing Scale (PCS)	1.1%	3.5%	1.1%	94.2%
Patients Global Perceived Rating of Improvement or Satisfaction	3.5%	1.3%	0%	95.2%
"Other" PROM scale (Orebro)	4.6%	0%	0%	95.4%
SF-36/SF12	0%	3.4%	1.1%	95.5%
Whiplash Disability Questionnaire	2.2%	0%	2.3%	95.5%
WHO-Brief	1.1%	0%	1.2%	97.7%
Bournemouth Questionnaire (BQ)	1.2%	0%	0%	98.8%
Neck and Upper Limb Index (NULI)	0%	1.2%	0%	98.8%
Pain Distress Scale	1.1%	0%	0%	98.9%
Northwick Park Neck Pain (NPQ)	0%	0%	0%	100%
Rapid Upper Limb Assessment (RULA)	0%	0%	0%	100%
Work Limitation Scale	0%	0%	0%	100%
Work Distress Scale	0%	0%	0%	100%
Nottingham Health Profile (NHP)	0%	0%	0%	100%

Other PROMs: Orebro=4.6% routinely, 0%=regularly and rarely

4.3.3.2 Utilisation of PROMs

Nearly 70% (n=1469) of physiotherapists in the survey reported that they would not use any PBOMs when assessing and/or managing NS-NP. The PBOMs which UK physiotherapists reported that they would use most often in the assessment/management of patients with NS-NP are summarised in Table 4.7. Selection of ULD PBOMs was used to determine the utilisation of ULD PBOMs in the assessment and/or management of NS-NP. The selection of instruments was variable but most PBOMs were never used and the single dimensional range of motion scales such as Goniometric measurement of neck motion (18.7%), Quantitative sensory tests (QST) (8.1%) and Rating of segmental joint mobility (8.1%) were the most commonly used PBOMs. Pain threshold perception tests (i.e. pain algometry), muscles function scales (e.g. neck muscles strength tests and

neck muscles endurance tests) and functional performance tests (e.g. functional performance test and functional capacity assessment) were almost always never used.

Table 4.7: Utilisation of PBOMs 30.1% (N=632)

Measures	Utilisation			
	Routinely: >70% of cases	**Regularly: 11-70% of cases**	**Rarely: 1-10% of cases**	**Never: 0% of cases**
Goniometric Measure of Neck Motion	18.7%	1.1%	0%	80.2%
Quantitative Sensory Test (QST)	8.1%	3.3%	1.1%	87.4%
Rating of Segmental Joint Mobility	8.1%	3.4%	0%	88.4%
Neural dynamic testing	3.4%	5.8%	2.2%	88.6%
Neck Muscle Strength test	6.9%	3.5%	0%	89.6%
Neurological exam	6.8%	2.3%	0%	90.9%
Posture alignment measures	3.5%	4.5%	0%	92.0%
Proprioception test	1.1%	1.1%	4.7%	93.1%
Neck muscle endurance testing	2.3%	3.5%	0%	94.2%
Neck muscle stability testing	1.1%	2.3%	1.1%	95.4%
Pain Algometry	2.3%	1.1%	0%	96.6%
Upper extremity muscle strength/endurance	1.1%	1.1%	1.1%	96.6%
Movement Diagram	1.2%	1.1%	0%	97.7%
Functional performance tests	1.1%	1.1%	0%	97.8%
Inclinometer of Neck Motion	1.1%	0%	0%	98.9%
Functional capacity assessment	0%	0%	1.1%	98.9%
Other physical or functional measure/measures	0%	0%	0%	100%

4.3.3.3 Physiotherapists' reported reasons for utilising OMs

The reasons that UK physiotherapists cited for using OMs in the assessment and/or management of NS-NP are summarised in Table 4.8. Two-thirds 65.3% (1371) of physiotherapists in this survey reported reasons for using OMs when rehabilitating their patients with NS-NP. Amongst physiotherapists, the reasons for doing so were variable, with setting treatment goals being the most prevalent reason for using OMs. Communicating with patient, fulfilling charting/documentation and communicating with other healthcare professionals were also commonly reported reasons for using OMs. However, documentation requirement, research, marketing and other reasons were less commonly cited reasons for using OMs.

Table 4.8: Reasons for utilising OMs 65.3% (N=1371)

Reasons	Frequencies	Percentage of cases
Setting treatment goals	1106	81.2%
Communicating with patients	975	71.6%
Fulfilling charting/documentation	526	39.4%
Communicating with other healthcare professionals	508	37.3%
Medicolegal documentation requirement	268	19.7%
Research	157	11.5%
Other reasons	74	5.4%
Marketing	48	3.5%

Other Reasons: Assessment of Treatment Progress = 3, Audit of Service Efficacy = 23, Monitoring Progression = 8, Prognosis = 35, Requirement Commissioners = 1, Sub-Groups, Better Patient Understanding = 1, No Reasons Mentioned = 3.

4.3.3.4 Physiotherapists' reported reasons for not utilising OMs

One-third 34.7% (n=730) of physiotherapists in this survey reported that they would not use OMs when rehabilitating NS-NP. The reported reasons are diverse, as shown in Table 4.9. The most endorsed reasons for not using OMs were a lack of clear guidance about the suitability of available OMs and a lack of time. Lack of access to information/knowledge about OMs, there is no need to use OMs, lack of resources (e.g. expensive to purchase) and other reasons were less commonly cited reasons.

Table 4.9: Reasons for NOT utilising OMs 34.7% (n=730)

Reasons	Frequencies	Percentage of cases
Lack of clear guidance about suitability of available OMs	579	82.1%
Lack of time	549	77.9%
Lack of access to information/knowledge about OMs	103	14.6%
There is no need to use OMs	88	12.5%
Lack of resources (e.g. expensive to purchase)	15	2.1%
Other Reasons	3	0.4%

Other Reasons: I feel that outcome measures fail to reflect patients=1, I have not looked into validity of specific measures=1, I use them when I think an improvement will be=1.

4.3.4 Comparison between groups of physiotherapists

Although the main focus of this survey was to describe current UK physiotherapy practice in relation to the utilisation of management approaches and OMs when rehabilitating patients with NS-NP. Additional objectives (see Section 4.1.1) were to explore whether the utilisation of multimodal management approach packages and/or OMs were associated with any of the demographic characteristics (see Sections 4.3.4.1 and 4.3.4.2).

4.3.4.1 Utilisation of multimodal management approach packages

There were statistically significant association between the multimodal management approach packages used and the majority of demographic characteristics of the UK physiotherapists as shown in Table 4.10. Package 1 was significantly more likely to be utilised by female physiotherapists, those with a greater number of years of clinical practice, those who manage a smaller proportion of patients with NP and those with no special interest in NP. Physiotherapists with the least experience, those from Northern Ireland or Scotland, those managing the largest proportion of patients with NS-NP and those with a special interest in NP were significantly less likely to be delivering a multimodal management strategy.

Table 4.10: Utilisation of multimodal management approach packages (N=2101)

Variables	Multimodal "Package 1" 34.1% (N=716)	Multimodal "Package 2" 14% (N=294)	Multimodal "Package 3" 24.2% (N=508)	Multimodal "Not used" 27.7% (N=583)
Gender				
Male	28.4% (N=258)	21.6% (N=196)	22.1% (N=201)	27.9% (N=253)
Female	38.4% (N=458)	8.2% (N=98)	25.7% (N=307)	27.7% (N=330)
Setting				
Exclusively in the NHS	32.4% (N=326)	10.1% (N=102)	28.2% (N=283)	29.3% (N=294)
Exclusively in Non-NHS	35.7% (N=224)	14.8% (N=93)	21.5% (N=135)	28.0% (N=176)
Combination of NHS & Non-NHS	35.8% (N=166)	20.5% (N=95)	19.4% (N=90)	24.4% (N=113)
Other Setting	0.0% (N=0)	100.0% (N=4)	0.0% (N=0)	0.0% (N=0)
Years of Practice				
0-5 Years	24.6% (N=77)	13.7% (N=43)	19.5% (N=61)	42.2% (N=132)
6-10 Years	23.8% (N=221)	18.0% (N=167)	28.0% (N=260)	30.2% (N=280)
11-15 Years	32.4% (N=128)	10.4% (N=41)	24.3% (N=96)	32.9% (N=130)
15+ Years	62.4% (N=290)	9.2% (N=43)	19.6% (N=91)	8.8% (N=41)
Nation				
Scotland	32.0% (N=95)	8.8% (N=26)	23.6% (N=70)	35.7% (N=106)
Northern Ireland	25.6% (N=31)	22.3% (N=27)	16.5% (N=20)	35.5% (N=43)
Wales	19.3% (N=55)	6.0% (N=17)	60.0% (N=171)	14.7% (N=42)
England	38.3% (N=535)	16.0% (N=224)	17.7% (N=247)	28.0% (N=392)
Postgraduate Training (PGT)				
Without PGT	32.7% (N=352)	15.8% (N=170)	26.1% (N=281)	25.4% (N=274)
With PGT	35.5% (N=364)	12.1% (N=124)	22.2% (N=227)	30.2% (N=309)
Proportion of Patients with Neck Pain				
<25%	44.6% (N=393)	12.9% (N=114)	18.7% (N=165)	23.7% (N=209)
25-50%	27.1% (N=251)	17.3% (N=160)	27.4% (N=253)	28.2% (N=261)
>51%	24.4% (N=72)	6.8% (N=20)	30.5% (N=90)	38.3% (N=113)
Special Interest in Neck Pain (SI)				
'No' to SI	38.4% (N=525)	16.2% (N=222)	20.5% (N=280)	24.9% (N=340)
'Yes' to SI	26.0% (N=191)	9.8% (N=72)	31.1% (N=228)	33.1% (N=243)

Multimodal Package 1 = Therapeutic Exercise + Manual Therapy + Patient Education, Multimodal Package 2 = Therapeutic Exercise + Manual Therapy, Multimodal Package 3 = Therapeutic Exercise + Any Other Approach, OMs = Outcome Measures.

4.3.4.2 Utilisation of OMs

There were statistically significant associations between OMs utilisation and the majority of the demographic characteristics as shown in Table 4.11. OMs were significantly more likely to be used by male physiotherapists, those working in non-NHS and other settings, those with 6-15 years of clinical practice, those without postgraduate training and those with no special interest in NP. Physiotherapists with the least experience, female physiotherapists, those practicing in Wales or England, those working for the NHS, those with postgraduate training, those with a very small proportion or a very high proportion of patients with NP and those with a special interest in NP were significantly less likely to incorporate OMs when rehabilitating their patients with NS-NP.

Table 4.11: Utilisation of OMs (N=2101)

Variables	Respondent (N=2101)	Responded 'Yes' To Using OMs 65.3% (N=1371)	Responded 'No' To Using OMs 34.7% (N=730)
Gender			
Male	43.2% (N=908)	70.0% (N=636)	30.0% (N=272)
Female	56.8% (N=1193)	61.6% (N=735)	38.4% (N=458)
Setting			
Exclusively in the NHS	47.8% (N=1005)	58.9% (N=592)	41.1% (N=413)
Exclusively in Non-NHS	29.9% (N=628)	72.9% (N=458)	27.1% (N=170)
Combination of NHS & Non-NHS	22.1% (N=464)	68.3% (N=317)	31.7% (N=147)
Other Setting	0.2% (N=4)	100.0% (N=4)	0.0% (N=0)
Years of Practice			
0-5 Years	14.9% (N=313)	60.5 % (N=190)	39.5 % (N=123)
6-10 Years	44.2% (N=928)	70.9% (N=658)	29.1% (N=270)
11-15 Years	18.8% (N=395)	44.3% (N=175)	55.7% (N=220)
15+ Years	22.1% (N=465)	74.8% (N=348)	25.2% (N=117)
Nation			
Scotland	14.1% (N=297)	57.2% (N=170)	42.8% (N=127)
Northern Ireland	5.8% (N=121)	40.5% (N=49)	59.5% (N=72)
Wales	13.6% (N=285)	41.1% (N=117)	58.9% (N=168)
England	66.5% (N=1398)	74.0% (N=1035)	26.0% (N=363)
Postgraduate Training (PGT)			
Without PGT	51.3% (N=1077)	67.0% (N=722)	33.0% (N=355)
With PGT	48.7% (N=1024)	63.4% (N=649)	36.6% (N=375)
Proportion of Patients with Neck Pain			
<25%	41.9% (N=881)	72.1% (N=635)	27.9% (N=246)
25-50%	44.1% (N=925)	63.2% (N=585)	36.8% (N=340)
≥51%	14% (N=295)	51.2% (N=151)	48.8% (N=144)
Special Interest (SI) in Neck Pain			
'No' to SI	65.1% (N=1367)	71.8% (N=982)	28.2% (N=385)
'Yes' to SI	34.9% (N=734)	53.0% (N=389)	47.0% (N=345)

4.4 Discussion

This national survey has captured information in relation to UK physiotherapists' current clinical measurement and management of patients with NS-NP and has provided data regarding their utilisation of multimodal management and measurement approaches. More specifically, the survey has led to insight into the current measurement and management of ULD in patients with NS-NP. Consequently, the aim and objectives of this investigation have been met. A total of 2101 responses were achieved from a wide range of physiotherapists practicing in the four nations of the UK. Participants had a varied length of experience and postgraduate training level and worked in a range of healthcare settings. The reminder of this section will present:

- Summary and discussion of the main findings
- Strengths of the study
- Limitations of the study
- Clinical implications
- Research implications
- Conclusion

4.4.1 Summary and discussion of the main findings

4.4.1.1 Management approaches

The survey findings indicate that a wide range of management approaches are currently used in UK physiotherapy practice for patients with NS-NP. However, the findings also indicate that the majority of participants were adopting an evidence-based approach to the management of patients with NS-NP. It appears that in line with the findings of recent systematic reviews and the recommendations of current guideline, the majority of physiotherapists in the UK reported using management approaches that are supported by strong evidence of effectiveness (Child et al. 2008, Kay et al. 2012, Bertozzi et al. 2013, O'Riordan et al. 2014). The multimodal management approach packages that incorporate a combination of therapeutic exercise and manual therapy with/without patient education programmes, delivered concurrently, are utilised to a high extent when rehabilitating patients with NS-NP. This is consistent with the current evidence-base that advocates the use of a multimodal management approach in physiotherapy rehabilitation (see Chapter 3 section 3.3). Further, the largest volume of neck related clinical management evidence recommended the use of a multimodal management approach and considered that a combination of exercise, manual therapy and patient

education is the optimal strategy to be used with this patient group (Gross et al. 2002, Kay et al. 2005, Gross et al. 2007, Hurwitz et al. 2008).

Therapeutic exercise was the most commonly used intervention as more than 80% of physiotherapists in this survey reported that they would use exercise when rehabilitating NS-NP. The majority of those who were utilising therapeutic exercise favoured exercise components (such as strengthening, stretching and general aerobic exercise), and this also is in a substantial agreement with the current guidelines which reported that therapeutic exercise is a fundamental component of physiotherapy rehabilitation of patients with NS-NP (Child et al. 2008). In addition, recent systematic reviews found moderate to strong evidence to support the use of specific exercise components (e.g. strengthening, stretching and general aerobic/strengthening/endurance exercise) alongside manipulation and/or mobilisation with some form of patient education for the short-and long-term benefits on pain, function and patient satisfaction for NS-NP (Kay et al. 2005, Gross et al. 2007, O'Leary et al. 2007, Ylinen et al. 2007, Hurwitz et al. 2008, Kay et al. 2012, Bertozzi et al. 2013, O'Riordan et al. 2014, Fredin and Loras 2017). It is also encouraging that management approaches such as electrotherapy methods, hydrotherapy, Feldenkrais and traction, which have limited or no evidence of effectiveness for patients NS-NP, were rarely or never utilised by UK physiotherapists when rehabilitating patients with NS-NP (Gross et al. 2007, Hurwitz et al. 2008).

However, one-third of UK physiotherapists in this survey appeared to be utilising lone management approaches that are not supported by strong evidence of effectiveness, such as massage therapy or acupuncture, or are utilising a multimodal approach package that is sub-optimal (e.g. exercise with any other intervention) (Trinh et al. 2006, Haraldsson et al. 2006, Ezzo et al. 2007, Gross et al. 2007, Vernon et al. 2007). In addition, the majority (60%) of physiotherapists in this survey reported that they do not consider upper limb rehabilitation when managing patients with NS-NP, and this seems inconsistent with the current evidence-base which suggests that there is a strong relationship between NS-NP and ULD. Given that the presence of NS-NP is a potential risk factor for the development and progression of ULD, and upper limb dysfunction may lead to NS-NP becoming recurrent, persistent or disabling (Eriksen et al. 1999, Daffner et al. 2003, Falla et al. 2004, Frank et al. 2005, Bot et al. 2005). Several studies point to the potential importance of incorporating upper limb rehabilitation strategies during the management of patients with NS-NP (McLean et a. 2011, Osborn and Jull

2013). For example, Osborn and Jull (2013) conducted a cross-sectional survey of patients with NS-NP (n=193) presenting for physiotherapy rehabilitation in an Australian general community. The purpose of the study was to establish the proportion of NS-NP patients who concurrently reporting ULD. Patient aged between 18 – 70 years and currently experiencing NP were included, but patients with cervical radiculopathy (clinical neurological signs) as well as neck or upper limb pathology were excluded. The study identified that 80% of patients with NS-NP reported ULD with one or more upper limb tasks because of the NS-NP. The study also found that there is a correlation between the severity of NS-NP and the level of ULD, in which higher severity of NS-NP was associated with higher ULD. The study concluded that the majority of patients with NS-NP often reported ULD, and this suggests that physiotherapists and clinicians involved in the management of patients with NS-NP should carefully evaluate the upper limb functional capacity using suitable OMs while assessing NS-NP to identify and quantify any ULD and include ULD rehabilitation in the management plan, if indicated (Osborn and Jull 2013).

There is no comparable study that has investigated UK physiotherapy practice regarding the management of patients with NS-NP to assist in the interpretation of this survey's findings. However, one recent international multi-professional survey which included physiotherapists (38%) was found (Carlesso et al. 2014). This cross-sectional study surveyed 360 clinicians from 17 countries, including the UK, to determine the practice patterns of clinicians involved in the management of patients with NP. Similar to the current survey, the findings indicate that exercise was the most frequently used intervention by physiotherapists and chiropractors for the management of NP. Furthermore, the findings also indicate that management approaches with low or very low evidence of effectiveness including traction were not being used. Another survey which investigated current practices of physiotherapists working in Swedish Primary Care and involved in the management of patients with low back pain, NP and sub-acromial pain was also found (Bernhardsson et al. 2015). This study validated and used a web-based questionnaire to survey 419 physiotherapists working in Primary Care in Western Sweden. Similar to the current survey, the findings indicated that exercise and patient education (advice) were found to be the most commonly used interventions when rehabilitating patients with NP. In addition, interventions with limited or no evidence of effectiveness such as electrotherapy and acupuncture were used to a great extent, as compared to the research carried out as part of this study (Trinh et al. 2006,

Kroeling et al. 2009). However, the survey undertaken as part of this research programme differs from other surveys because it captured information on the utilisation patterns of the multimodal management approach packages and the level of ULD rehabilitation during the management of patients with NS-NP.

In summary, this research found that around 40% of physiotherapists are undertaking optimal evidence-based practice in the management of patients with NS-NP. Despite this, there is considerable room for optimising the management of NS-NP, by increasing the use of multimodal management packages of treatment, and by incorporating ULD rehabilitation where indicated for patients with NS-NP.

4.4.1.2 Utilisation of OMs

The survey findings indicate that OMs are poorly incorporated by UK physiotherapists in the assessment and during the management of patients with NS-NP. Physiotherapists in the UK appeared to be either not utilising OMs or utilising inappropriate OMs when evaluating their patients with NS-NP. A third of physiotherapists in this survey reported that they never utilise OMs when evaluating patients with NS-NP. The most commonly reported reasons for not utilising OMs were a lack of clear guidance about the suitability of available OMs and a lack of time. This is inconsistent with the clinical guidelines and professional bodies recommendations regarding the utilisation of OMs. Clinical guidelines and professional bodies suggest that routine utilisation of standardised OMs is a fundamental part of physiotherapy rehabilitation and considered to be the optimal way to facilitate evidence-based practice (Hammond 2000, Rudd et al. 2000, CSP 2005, College of Occupational Therapists 2007, American Occupational Therapy Association 2010). In addition, UK organisations such as the HCPC, CSP and the NHS explicitly recommend routine utilisation of standardised OMs wherever practicable (NHS 2010, CSP 2012, HCPC 2013). Meanwhile, in standard 12 of the Standards of Proficiency for Physiotherapists (HCPC 2013), the HCPC suggest that physiotherapists must be able to collect and document qualitative and quantitative data on their patient's condition by using standardised OMs. This is to assure the quality of clinical practice by meeting the patient's needs and changes in health, demonstrating the significance of physiotherapy by enabling physiotherapists to prove their impact and cost-effectiveness (HCPC 2013).

Routine utilisation of standardised OMs is important since it can be used for a variety of purposes in clinical practice. First, before an intervention for screening of symptoms/function for diagnosis and prognosis purposes (Lansky et al. 1992), classifying patients into meaningful sub-groups and setting treatment goals (Kramer and Holthaus 2006). Second, during an intervention to monitor condition progression, detect changes in pain and disability and facilitate communication between physiotherapists and patients and other healthcare professionals (Garland et al. 2003, Bot et al. 2007, Nordin et al. 2008). Finally, after an intervention, they can be used to determine the effectiveness, efficiency and cost-effectiveness of the intervention (CSP 2012).

The findings also indicate that nearly all physiotherapists in this survey reported that they do not utilise any OM to evaluate the upper limb functional capacity when assessing/managing patients with NS-NP. Given the relationship between the presence of NP and the presence of ULD (McLean et al. 2011, Osborn and Jull 2013), indicating that many patients with NS-NP are likely to present with associate ULD, the author strongly recommend that physiotherapists should routinely evaluate upper limb functional capacity using suitable OMs in the assessment and during the management of patients with NS-NP. This is to identify and quantify any ULD and create a rationale for including upper limb rehabilitation in the management plan, if indicated.

The majority of the two-thirds of UK physiotherapists who were utilising OMs reported that they would consistently use single dimensional numeric pain and range of motion rating scales such as the Visual Analogue Scale (VAS) or Numeric Rating Scale (NRS) and Goniometric measurement of neck motion. Single dimensional scales such as pain and range of motion are narrow parameters of NS-NP and cannot capture information in relation to constructs such as physical function, psychological, social capacity and quality of life (Mintken et al. 2009). However, these constructs, including physical and functional limitations, psychological distress and reduced quality of life, that are often associated with NS-NP, were rarely or never measured by UK physiotherapists when evaluating patients with NS-NP. Limited utilisation of OMs to measure valid constructs in the assessment and during the management of patients with NS-NP is inconsistent with the evidence-based and may contribute to inadequate evaluation of patients with NS-NP (Borghouts et al. 1998, Hoving et al. 2004, Bot et al. 2005, Binder 2007, Haldeman et al. 2008) and may be one of the factors that contribute to suboptimal management of NS-NP (McLean et al. 2011, Osborn and Jull 2013).

There is no comparable study that investigated UK physiotherapy practice regarding OMs utilisation in the assessment and during the management of NS-NP to assist in the interpretation of this survey's findings. However, Jetta et al. (2009) investigated the extent of current utilisation of standardised OMs as well as the perceptions of physical therapists regarding the benefits and barriers to OMs use. They surveyed 1000 physical therapists who were randomly selected from a list of all members of the American Physical Therapists Association (APTA). Similar to the current survey, a substantial proportion of respondents reported not utilising OMs when managing patients with musculoskeletal conditions and lack of time as well as confusion regarding the selection of OM were reported as the main barriers to OMs utilisation (Jette et al. 2009). Another survey, investigating OMs utilisation in the management of NP in various disciplines such as chiropractors, manual therapists, massage therapists, physicians and physiotherapists, was also found (MacDermid et al. 2013). This international survey recruited 381 clinicians, of which physiotherapists comprised 32% of respondents, who completed an online questionnaire. Similar to this study, the findings established the poor utilisation of standardised OMs across all the included disciplines in the management of patients with NP. However, the survey undertaken as part of this research differs from these surveys, in that it offers additional insight on the current evaluation of upper limb functional capacity in the assessment and during the management of patients with NS-NP as well as identifying those UK physiotherapists who were utilising a multimodal measurement approach that combines the utilisation of PROMs and PBOMs.

4.4.2 Strengths of the survey

This study robustly developed a survey instrument designed in accordance with the Web-Based Survey Design Standard (Crawford et al. 2005). The design (web-based) incorporated advanced and iterative features, provided a dynamic survey process which facilitated the simplicity and brevity of the survey instrument, and this in turn contributed to achieving fast and accurate low-cost data collection and data analysis along with a large sample size. An innovative online methodology that was used in the recruitment of participants and administration of the survey instrument potentially contributed to the large sample size. The large sample size achieved (4.3% of the whole UK physiotherapy population), which is broadly comparable in demographic characteristics to the total UK physiotherapy population, points to the validity of these

findings (Morgan and Harmon 1999). This is the first study to describe current UK physiotherapy practice with regard to the management and measurement of NS-NP. In addition, this is the first study, in the UK or internationally, to investigate the current level of measurement or management of ULD in patients with NS-NP.

4.4.3 Limitation of the survey

This web-based survey has several possible limitations. This survey will not have gained responses from physiotherapists who do not have access to or do not use the internet. However, the internet has become a fundamental vehicle in the navigation of everyday life (Hughes et al. 2012), and the majority of adults living in the UK now have access to the internet according to the UK office for National Statistics annual report for 2016 (Office for National Statistics 2016). In addition, evidence suggests that a substantial component of the UK healthcare professionals including physiotherapists are regularly using popular Social Networking Sites such as Facebook, Tweeter, LinkedIn and YouTube for social interaction and information exchange as well as business advertising (Ahmed et al. 2012). The risk of missing out physiotherapists who do not have internet access may have been present but it is considered to be of low risk.

A simple random sampling procedure (probability sampling), in which each population person/member has a known non-zero chance of being selected for inclusion in the sample, is considered to be the optimal approach to produce a representative survey from which the findings can be generalised to the wider population (Morgan and Harmon 1999). This was not possible in this survey. At present, the necessary list (sample frame) of UK physiotherapists is not available. From the sampling methods used in this survey (non-probability) there was no reliable way to determine potential participants, estimate the sample size and/or estimate the true response rate. Consequently, there was no way to identify non-respondents or to assess whether the respondents were different to the national physiotherapy population. Consequently, non-respondent bias may be present (Bosniak et al. 2005). However, a large sample size (4.3% of the whole UK physiotherapy population) was obtained, diversity between participants was achieved and the survey population was similar on many demographic characteristics to the UK population of physiotherapists (see Section 4.3.1.1). Consequently, it is likely that the findings are generalisable to the UK physiotherapists who manage patients with NS-NP.

Measuring physiotherapists' current practice was self-reported clinical behaviour based on direct questions for a clearly defined condition (NS-NP). Consequently, social desirability bias may be present, since we cannot be completely confident how items were interpreted and whether the management approaches and OMs instrument names were interpreted the same way across respondents from different UK nations, working in many different settings and have variety of training levels. Within practice, utilisation of management approaches or OMs may therefore be different. However, the survey instrument was extensively piloted by clinicians in order to try to ensure that items were easy to understand and interpret by clinicians. The survey findings provide a rough indicator of practice in the UK and may not reflect nuanced practice or decision-making process that inform practice. It is possible that physiotherapists may have wanted further information in order to make more informed clinical decisions regarding the measurement and management of their patients with NS-NP. However, more in-depth qualitative studies would be required to support a survey of this nature and to make those determinations.

Finally, psychological and relaxation interventions were identified by some respondents as "other" interventions. It is possible that more respondents might have reported utilising these interventions if they had been incorporated as standard items on the survey instrument. Consequently, utilisation of these "other" interventions may be under-reported.

4.4.4 Clinical implications

4.4.4.1 Management

This survey demonstrated that the majority of physiotherapists in the UK are utilising management approaches that are supported by strong evidence of effectiveness, consistently utilise active approaches and the multimodal management approach packages 1 and 2. However, a third of physiotherapists in the UK are utilising either a lone intervention which is not supported by strong evidence of effectiveness or multimodal management approach package 3 which is sub-optimal. The reasons of this are likely to be multifactorial, including patient and physiotherapist preferences and interactions regarding the selection of interventions (Child et al. 2008). Physiotherapists' preferences for the management of NS-NP were found to be primarily influenced by the level and place of their training, the setting and the type of working place, availability of resources, special interest and possibly the stage of healing. Meanwhile patients'

preferences were found to be primarily influenced by their previous experiences (Tsakitzidis et al. 2013, Carlesso et al. 2014). Updating and disseminating clinical guidelines may facilitate clinician education on how to reset a patient's preferences, increase the use of multimodal approach packages 1 and 2 and reduce the use of interventions that have been shown to be ineffective.

Two-thirds of physiotherapists in the UK do not incorporate upper limb rehabilitation strategies in the management of patients with NS-NP. The reasons for this may be related to how physiotherapists evaluate the upper limb functional capacity during the assessment of patients with NS-NP. Simple screening of shoulder range of motion is often recommended and used by UK physiotherapists to rule in/out the presence of co-existing shoulder or upper limb dysfunction (Petty 2011), and this is insufficient since range of motion does not correlate conclusively with disability (Olson et al., 2000; Poitras et al., 2000; Kwak et al., 2005). Failure to sufficiently evaluate the upper limb functional capacity in the assessment of patients with NS-NP could lead to the development and progression of ULD which may contribute to NS-NP become a recurrent, persistent or disabling condition. This process may contribute to poor treatment outcomes and reduced quality of life for patients with NS-NP. Clinicians should give careful consideration as to how best to evaluate upper limb functional capacity using standardised measures and include upper limb rehabilitation, if indicated.

4.4.4.2 Measurement

The findings of this survey, which are consistent with the findings of other recent comparative surveys (Jette et al. 2009, MacDermid et al. 2013), established the current poor utilisation of OMs in the assessment and during the management of patients with NS-NP. This suggests that physiotherapists in the UK appear to be inadequately evaluating their patients with NS-NP. Inadequate evaluation of patients with NS-NP before an intervention may result in failing to recognise deficits that would classify patients into meaningful subgroups and facilitate the clinical reasoning process which may in turn lead to the development of the most appropriate management plan. Meanwhile, failure to make ongoing evaluations during the management process could be a barrier to evaluating the impact, effectiveness and efficiency of the given intervention, which may also contribute to poor treatment outcomes and reduced quality of life for patients with NS-NP.

Standardised OMs that evaluate valid and relevant constructs for NS-NP including physical and functional limitations, psychological distress, and reduced quality of life were rarely or never utilised; instead single dimensional impairment marking scales such as VAS, NRS and the goniometric measure of neck motion were frequently utilised in the assessment and during the management of patients with NS-NP. This could be because these scales are generic, easy to use, quick to administer and interpret (verbally), and therefore may be seen as feasible scales for use in busy clinical practice (Mintken et al., 2009). However, this is insufficient since these scales cannot capture information in relation to the aforementioned valid and relevant constructs and reporting bias is often present when administered by the clinician (MacDermid et al. 2013). Furthermore, these scales have less reliability and lacked responsiveness to change in the patient's condition (Mintken at al. 2009). Another reason for this could be that physiotherapists in the UK do not differentiate between impairment scales and OMs.

In this survey, pragmatic reasons such as a lack of clear guidance regarding the availability of suitable OMs and a lack of time were frequently found to impede the utilisation of OMs. The reason for this could be because there are a wide variety of OMs available and that clinical guidelines and professional bodies which explicitly recommended and advocated routine utilisation of OMs in clinical practice rarely specify which OM should be used, and this causes uncertainty among physiotherapists (Connell and Tyson 2012). In addition, the majority of the available standardised OMs are patient-reported questionnaires that require resources (e.g. time, pen and papers) and proficiency in English for completion, making them impractical for busy clinical practice for majority of physiotherapists. Therefore, it is important to identify OMs that are easy for patients and clinicians to use and interpret.

Upper limb functional capacity is poorly evaluated in the assessment and during the management of patients with NS-NP. Failure to adequately evaluate the upper limb functional capacity by using suitable upper limb OMs often lead to the development and progression of ULD. ULD could have a detrimental cyclical effect on the neck and may contribute to chronic and persistent neck and upper limb problems. This may lead to reduced quality of life for patients with NS-NP.

4.4.5 Research implication

4.4.5.1 Management

The findings of this survey highlighted that there is a gap between evidence-based and current practice when evaluating and rehabilitating patients with NS-NP, and this suggests the need for better evidence-based dissemination and knowledge translation. Further research is required to address how to promote the use of a multimodal management approach for patients with NS-NP, and how to increase physiotherapists' awareness regarding the importance of including upper limb rehabilitation during the management of patients with NS-NP. Investigation of the effectiveness of treatment approaches, for which evidence is limited or conflicting but still used in practice, may be also required. Further information is also needed regarding the management preferences and treatment choices for physiotherapists and patients alongside strategies to match intervention preference with the evidence-based practice (Sackett et al. 1996, Sackett et al. 1997). This would facilitate appropriate allocation of healthcare resources and minimise expenditure for ineffective interventions.

4.4.5.2 Measurement

Despite the importance of evaluating patients with NS-NP, physiotherapists in the UK are still insufficiently evaluating their patients with NS-NP. Pragmatic reasons often impede the utilisation of OMs as well as the lack of clinical OMs. This highlights the need for an efficient OM collection system which ensures the successful incorporation of standardised OMs. Further, this system should provide clear guidance on the choice of OMs and remove barriers such as lack of knowledge and confidence in selecting and utilising standardised OMs. This should facilitate to overcome the complexity of establishing a culture of routine data collection using standardised OMs.

Further research is also required to provide a valid and reliable clinical ULD OM that will support the recommended assessment and management of patients with NS-NP. Such a measure, which can accurately examine upper limb functional capacity in the assessment stage and monitor the progress of patients during the rehabilitation programme, will enable physiotherapists and clinicians involved in managing patients with NS-NP to deliver safe, effective, and efficient treatment for this patient group.

4.5 Conclusion

This chapter has described current physiotherapy practice regarding the management and measurement of patients with NS-NP. It has demonstrated that the most frequently reported management approaches for NS-NP are those that also have strong evidence for their effectiveness. It has also indicated variable use of management approaches with low or unclear evidence of efficacy. This suggests the urgent need for updated clinical guidelines to support physiotherapists to reduce the use of ineffective management approaches and prompt the use of the multimodal management approach. The results of section 4.3.3.1 of this survey (OMs utilisation) established the poor utilisation of OMs as well as the limited evaluation and rehabilitation of ULD in patients with NS-NP. This suggests that further research is needed to establish a core outcome set, and outcome measures that are standardised, valid and suitable for use in clinical practice. Innovative strategies are also needed to prompt the implementation of OMs in clinical practice. This might further support targeted, tailored interventions for patients with NS-NP.

The findings from this chapter together with the findings from chapter 2 (the systematic review) of this thesis have underlined the gap in research regarding ULD OMs that are suitable, standardised and adequately validated for use in clinical practice and research practice for patients with NP. In addition, it justifies the further development of the Single Arm Military Press (SAMP) test, which is performance-based, brief, easy to administer, score and interpret, and therefore has the potential to be useful in clinical practice to accurately examine the upper limb functional capacity and monitor the progress of patients during their rehabilitation programmes. This should facilitate to reduce the potential for ULD to have detrimental cyclical effect on the neck by undertaking early upper limb rehabilitation, and this will enable physiotherapists to deliver safe, effective and efficient treatment for patients with NS-NP.

The next chapter is a validation study which addresses the third research question (see Section 1.5) and describes the evaluation of the acceptability and feasibility of the SAMP test at lower weights (½-kg, 1-kg, 1½-kg) in a cohort of Egyptian female patients with NS-NP.

Chapter 5: Measuring upper limb disability in a neck pain population: Evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test.

5.1 Introduction

This chapter describes a study that explored the acceptability and feasibility of the Single Arm Military Press (SAMP) test from both the patient and clinicians' perspective. It is the first stage of research that evaluate the measurement and practical properties of this instrument and addresses the third question of this thesis (see Section 1.5).

The UK national survey of neck pain (described in Chapter 4 of this thesis) provided empirical evidence that the majority of the UK musculoskeletal physiotherapists do not include upper limb rehabilitation strategies when managing patients with non-specific neck pain (NS-NP). The survey also established that nearly all UK musculoskeletal physiotherapists reported not using any upper limb OMs when assessing and/or managing patients with NS-NP, and the most frequently reported reasons for this were a lack of clear guidance regarding the suitability of available OMs and a lack of time. This indicates the specific need for clinically suitable UL OMs that are simple, quick, inexpensive and easy to administer and interpret.

The systematic review on the measurement and practical properties of OMs that were developed or validated to measure upper limb disability (ULD) in patients with NS-NP (see Chapter 2) highlighted the lack of good quality evidence for any of the identified OMs. Synthesis of the results suggested that the SAMP test is a promising ULD OM for patients with NS-NP (see section 2.4). Given that it is a performance-based test, it has the theoretical advantages of better reliability, greater sensitivity to change and low vulnerability to external variance, such as culture, cognition, language and level of education (Latham et al. 2008, de Vet et al. 2011). The SAMP test is also quick, inexpensive and easy to use; however, further validation in good quality studies is urgently required to improve its utility for clinical practice and research.

The SAMP test has undergone a series of preliminary investigations, in which a 3-kg hand weight was used in its practical application (McLean et al. 2010a). However, anecdotal evidence (personal communication of the developer of the SAMP test and the Director of Studies, Dr Sionnadh McLean, for this PhD) suggested that a 3-kg weight

may be unsuitable (too heavy) and therefore unethical for patient populations with NP. Therefore, this study was undertaken to explore the acceptability and feasibility of the SAMP test at lower weights on patient populations with NS-NP.

5.1.1 Aim

The overarching aim of this study was to explore the acceptability and feasibility of the SAMP test using lower weights (½-kg, 1-kg, 1½-kg) in female patients with NS-NP. The objectives were to:

- Explore the acceptability of the SAMP weight in female patients.
- Explore patient willingness and ability to perform the SAMP test's tasks despite their neck and upper limb symptoms.
- Explore patient burden regarding the time and effort required for the SAMP test performance.
- Explore the acceptability of the SAMP test in relation to its overall administration and completion for the examiners
- Explore the burden for examiners in relation to the time and resources required when incorporating the SAMP test.

5.1.2 Acceptability and feasibility: Concepts

These properties have been overlooked by the vast majority of the literature and less frequently examined. However, Fitzpatrick et al. (1998) highlighted acceptability and feasibility as essential practical properties that should be possessed by all OMs. There is no consensus about the definitions of these terms and frequent overlap in the definition occur (e.g. acceptability and burden, feasibility and burden for patient and clinician). This suggests the need for a study similar to the COSMIN taxonomy (Mokkink et al. 2010b) for these properties.

5.1.2.1 Acceptability

OMs need to be acceptable to patients in order to help to eliminate avoidable distress to those already coping with pain and/or disability, and to obtain a measurement score (Fitzpatrick et al. 1998). Therefore, it is necessary and should be established prior to other measurement properties such as reliability and validity (Selby and Robertson 1987). Acceptability is defined as the ability and willingness of a patient from the target population to complete questions or tasks related to an OM (Fitzpatrick et al. 1998). It is also described as the difficulties a clinician may encounter during the

administration and/or interpretation of an OM (e.g. the length and complexity of the overall testing procedure) (Fitzpatrick et al. 1998). Direct assessment of acceptability, by obtaining both the patient and clinician's opinion about the OM under study immediately after administration using interviews, is considered to be an optimal parameter of acceptability (Sprangers et al. 1993). The interview should include questions about whether the new OM was difficult, confusing, annoying, upsetting, distressing or whether items/tasks should be removed.

5.1.2.2 Feasibility

In addition to patient and clinician acceptability, it is essential to evaluate the impact and burden upon the patient and clinician when administrating an OM (Lansky et al. 1992). Feasibility is described as the time, training and effort required from a clinician to measure patient outcomes using an OM (Erikson et al. 1995). The resources required (e.g. purchasing, extra staff, or extra training) and the time needed for the administration procedure are considered to be optimal parameters of feasibility (Fitzpatrick et al. 1998). This suggests that free access, brevity, simplicity in administrating, scoring and interpreting the score of an OM indicates greater feasibility (Read et al., 1987; Feeny and Torrance, 1989; Nelson et al., 1990).

5.2 Methods

5.2.1 Study design

A pragmatic randomised controlled validation study was designed to explore the acceptability and feasibility of the SAMP test at lower weight (½-kg, 1-kg, 1½-kg) on female patients with NS-NP. This study was conducted, analysed and interpreted in accordance with the **I**nternational **S**ociety for **Q**uality **O**f **L**ife research (ISOQOL), which provides a minimum requirement checklist for the evaluation of the measurement and practical properties of an outcome measure (Reeve et al. 2013). A total of 70 Egyptian female patients with NS-NP were randomly allocated into one of three testing groups. The stratification of patients was according to the severity of their neck and upper limb symptoms using the NDI scores; this was done to ensure balance between the three groups on important criteria. Patients were recruited from the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital (Egypt). This procedure was carried out on six days between 30th November 2015 and 26th December 2015.

5.2.2 Ethical consideration

This study was approved by Tanta Universal Teaching Hospital (Rheumatology and Physical Therapy Department) (see Appendix 7). Subsequently, approval was sought and gained from the Health & Wellbeing Faculty Ethics Committee at Sheffield Hallam University to conduct this study in Egypt (SHUREC) (see Appendix 8).

5.2.3 Study setting

This study took place in the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital, which is a new Universal Teaching Hospital located in the El-Gharbia province. This hospital serves over 1.5 million people every year via self-referral from four different provinces in the heart of Egypt. Anecdotal reports suggest that approximately 10% of these 1.5 million people are patients with NP.

5.2.4 Training and information delivered to staff

Twenty-four hours prior to the face-to-face assessment of patient participants and SAMP testing, clinical staff who were involved in the data collection in this study attended a 30-minutes practical training and information giving session at the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital (Egypt). The session was delivered by the researcher (AA)and covered the purpose of the study; a brief outline of the SAMP test description and practical application; standardised demonstration of the warm-up; standard utilisation of the SAMP test's technique and how to recognise compensatory strategies that lead to ineligible lifts; and the SAMP scoring system.

5.2.5 Participants recruitment

The manager of the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital was approached and details of the study were explained to him. He agreed to host the study in the Rheumatology and Physical Therapy Department and also agreed to be involved alongside three clinical staff (physicians) from the same department. Subsequently, the researcher was provided with a list of 80 patients who had visited the Rheumatology and Physical Therapy Medicine Department or were on the waiting list with a diagnosis indicating neck pain of non-specific origin.

Ethics protocols usually insist that potential research participants in the UK are invited to consider participating in a research study by initially mailing them a letter containing

an invitation to participate in the study and requesting that interested participants contact a member of the research team. However, in Egypt, where this study was conducted, the mail services are unreliable and time consuming since houses in many provinces are identified by the nearest well-known shop or building rather than numbers. Furthermore, the obtained list included a patient's name, gender, mobile number and/or landline telephone numbers as well as the area where this patient resided (that is, no accurate address was available). Therefore, potentially eligible patients were telephoned by the researcher who explained the aim of the study and conducted a phone screening to confirm their provisional eligibility, gain verbal consent and invite them to attend a single face-to-face assessment and testing session. A telephone checklist of clinical and demographic questions was completed for each patient to ensure standardisation of the information given and to cover the inclusion and exclusion criteria (see Appendix 9). If provisional eligibility was confirmed and verbal consent obtained, an appointment for the face-to-face assessment and SAMP testing was organised. In addition, a convenient method to send the patient information sheet was agreed (see Appendix 10). A private company was commissioned to deliver a hard copy of the patient information sheet to each verbally consented patient at least 48 hours before their assessment and testing session. Patients were asked to carefully read the study information sheet and discuss potential participation with their family and friends. Patients were also informed that participation in the study was entirely voluntary and that they could withdraw from the study at any time.

5.2.5.1 Inclusion criteria

Female patients were considered for inclusion in this study if they were: (1) an adult ≥ 18 year of age, (2) attending or referred for physiotherapy treatment at Tanta Universal Teaching Hospital (Egypt), (3) experiencing NS-NP with/without referred symptoms into the head or upper limb, and (4) scoring at least 10 (out of 100) in the Neck and Disability Index (NDI) questionnaire. NS-NP is defined here as "pain perceived as arising from anywhere within the region bounded superiorly by the superior nuchal line, inferiorly by the transverse line through the tip of the first thoracic spinous process, and laterally by the sagittal planes tangential to the lateral border of the neck (Merskey and Bogduk 1994), and that the pain was not caused by any serious acute trauma (e.g. 'Whiplash Association Disorder), or neurological disorder (e.g. Cervical Radiculopathy, Nerve Root compression).

5.2.5.2 Exclusion criteria

Patients were excluded from this study if they: (1) had a neck condition that required urgent treatment, (2) had any potentially serious condition systemic disease, progressive or worsening neurological disorders, inflammatory conditions or major trauma, (3) had previous traumatic injury to the UL/shoulder girdle, (4) were unable/unwilling to do physical tasks using their UL, or (5) were unwilling to complete questionnaires.

5.2.6 Allocation procedure

In order to ensure balance between the three testing groups and accurate proportional representation of the sample, at first, patients were stratified into four groups based on their NDI score, (low, moderate, severe, extremely severe) (Vernon 1991). The first group comprised 20 patients who scored between 10-29 out of 100, second group had 28 patients who scored between 30-49, third group had 17 patients who scored between 50-68 and the final group had 5 patients who scored 69-100. Randomisation procedure in each of these groups was then used to allocate patients into the final three testing groups. This led to randomly allocating 23 patients in the first group for SAMP testing using ½-kg weight, 24 patients in the second group for SAMP testing using 1kg weight and 23 patients in the final group for SAMP testing using 1½-kg weight.

5.2.7 SAMP test protocol

The SAMP test protocol was designed to address tasks of functional relevance, which challenge the UL (neck, shoulder elbow, arm and hand) and are typically impaired in patients with NS-NP (i.e. carrying, lifting, and repetitive overhead movement) (McLean et al. 2007, McLean et al. 2011, Osborn and Jull 2013). The SAMP test uses readily available and inexpensive equipment (one dumbbell), and it is very easy to score (number of repetitions within 30 seconds). The task evaluates the strength and endurance of the UL, with the expectation that the difficulty in sustaining overhead activity within 30 seconds would discriminate between NS-NP patients with varying degrees of UL functional limitations. The SAMP test is conducted with the patient in the standing position with their feet positioned at shoulder width. The patient is asked to carry a dumbbell and to lift it, using their dominant hand, to shoulder level (see Figure 5.1A). The patient is requested to raise their hand with the dumbbell directly overhead by extending through the elbow (see Figure 5.1B) and to repeat this process as fast and as frequently as possible for 30 seconds (McLean et al. 2010a).

Fig 5.1A

Fig 5.1B

Figure 5.1 SAMP Test Protocol

5.2.8 Study protocol

Patients who were found provisionally eligible and verbally consented to take part in the study were booked to attend a single assessment and SAMP testing session at Tanta Universal Teaching Hospital (Egypt). The session took up to 45-minutes, in which patients completed the NDI questionnaire (Arabic version) on their arrival (see Appendix 11), after which stratification based on the NDI score was executed. Patients were then invited to the face-to-face assessment with the researcher and/or the department manager in a designated room, in which the patient's weight and height were recorded, and a subjective examination was carried out using standardised clinical questions (see Appendix 12). Patients were then randomly allocated into one of the three testing groups and requested to meet immediately with the relevant examiner for the SAMP testing, if they were found to be eligible, happy to proceed and consented in writing (see Appendix 13). Each testing group in the study was led by one examiner in a designated room. The SAMP testing was done for each patient individually and started with a brief warm-up, which included shoulder shrugs and flexion exercises as well as range of movement exercises for the neck and UL (see Appendix 14). This was followed by an explanation/description and demonstration of the SAMP test procedure by the examiner (see Appendix 15). The patient was then instructed by the examiner to perform the SAMP test. At the end of the testing procedure, patients were directed back to the researcher to complete the data collection process (see Section 5.2.9 below).

5.2.9 Data collection

A total of three examiners, physicians, with at least 3-years of experience of working with musculoskeletal patients were involved in the data collection of this study. The department manager who is a professor and member of teaching staff in the faculty of medicine at Tanta University was involved in the face-to-face assessment of all patients. One member of staff (BH) collected the SAMP data from the 23 patients who were allocated to be tested using ½-kg dumbbell and formed group one; a second member of staff (GE) collected the SAMP data from the 24 patients who were allocated to be tested using the 1-kg dumbbell and formed group two; and a third member of staff (HA) collected the SAMP data from the final 23 patients who were allocated to be tested using the 1½-kg dumbbell and formed group three. All examiners completed the data collection sheet (see Appendix 16) for each patient in their group regarding the SAMP score (that is, the number of valid SAMP reputations within 30 seconds); administration time (that is, description, demonstration of the SAMP procedure by the examiner as well as instruction and the SAMP performance by patient); and the completion time which included the warm-up and the administration time. Patients who competed the SAMP testing were immediately directed back to the researcher who first, measured their neck and UL symptoms severity (immediately after testing) using a 0-10 scale of pain severity where 0 indicates no symptoms and 10 indicates the worst possible symptoms (see Appendix 17). Second, acceptability of the weight used in the SAMP testing in their group was measured using a 1-9 Likert scale, where 1 indicates extremely light weight and 9 indicates extremely heavy weight (see Appendix 18). Third, acceptability of the SAMP test instruction and performance was measured using a 1-9 Likert scale where 1 indicates extremely easy to understand and perform and 9 indicates extremely difficult to understand and perform (see Appendix 18). Fourth, acceptability regarding patients' ability to perform the physical tasks required in the SAMP test was measured using a 1-9 Likert scale where 1 indicates extremely easy to do and 9 indicates extremely difficult to do (see Appendix 18). Finally, the feasibility of the SAMP test was assessed from the patient perspective in relation to the time and effort required using a 1-9 Likert scale where 1 indicates highly suitable and 9 indicates highly unsuitable (see Appendix 18). Patients were then discharged and telephoned the following day by the researcher (ASEA) to monitor and measure the severity of their neck and upper limb symptoms (24 hours after testing) on the 0-10 scale of pain severity as mentioned above (see Appendix 17).

Following completion of the SAMP testing, the three examiners involved in the data collection were requested to record their qualitative input regarding the SAMP test procedure. They recorded their opinion regarding the SAMP test's explanation, demonstration and instruction that they provided to each patient as well as the overall administration and completion using a 1-9 Likert scale where 1 indicates extremely easy and 9 indicates extremely difficult (see Appendix 19). Examiners were also requested to record qualitative responses regarding the resources required (e.g. time and cost) when using the SAMP test using a 1-9 Likert scale where 1 indicates highly appropriate and 9 indicates highly inappropriate (see Appendix 19). Furthermore, examiners were asked about whether there was a need for: (1) extra training to understand the application of the SAMP test's procedure, (2) extra staff to support the application of the SAMP test's procedure, and/or (3) technological support to facilitate the application of the SAMP test procedure on a 1-5 Likert scale where 1 indicates strongly agree and 5 indicates strongly disagree (see Appendix 19).

5.2.10 Outcome assessment

The primary outcomes for this study were the acceptability and feasibility of the SAMP test procedure from both the patient and clinicians' perspective.

Patient acceptability was evaluated in terms of the difficulty of using the weight provided in the practical application of the SAMP test, as well as understanding the instruction to correctly perform the test. Feasibility (patient burden) was assessed in terms of the time and effort required to complete the SAMP test (Fitzpatrick et al. 1998). This was determined by assessing each patient's view using a structured qualitative exit feedback interview (Sprangers et al. 1993). The interview questions explored the patient's experience of the weight used (extremely light - extremely heavy), the difficulties of understanding the instruction (extremely easy - extremely difficult), and the time and effort required (highly suitable – completely unsuitable) (see Appendix 18). To ensure the accuracy and precision of the patient experience when assessing acceptability and feasibility, a Likert scale with nine response categories (1-9) was used for each question (Avis and Smith 1994). In addition, the Visual Analogue Scale (VAS) with eleven response categories (0-10) was used to measure the impact of the SAMP testing procedure on patients when using different weights, as this may influence their perception of acceptability and feasibility of the SAMP procedure (Remington et al. 1979). Both Likert and VAS scales have been extensively validated

and found to be reliable, valid and responsive in measuring symptom severity as well as participant satisfaction and acceptability (Bond and Lader 1974, Scott and Huskisson 1977, Remington et al. 1979, McCormack et al. 1988, Wewers and Lowe 1990, Jaeschke et al. 1990, Bowling 1995, Bolton and Wilkinson 1998, Vickers 1999, van Dijk et al. 2002). Meanwhile, the NDI questionnaire, which is a standardised OM for measuring disability due to NP, was used to measure the neck symptoms at the baseline to confirm eligibility and to stratify the patients into four groups according to the degree of their symptom severity.

Acceptability to clinicians was evaluated in terms of the difficulty clinicians encounter during the administration of the SAMP test, such as the length and complexity of the overall testing procedure (Fitzpatrick et al. 1998). Meanwhile, feasibility (clinician burden) was assessed in terms of the resources required (cost and time) and whether or not there was a need for extra training, extra staff or unusual experimental conditions when administering the SAMP test (Feeny and Torrance 1989, Aaronson 1992, Lansky et al. 1992, Erickson et al. 1995). This was established using a qualitative feedback interview for each examiner, in which questions explored the examiner's experience with the SAMP testing procedure; e.g. providing explanation, demonstration, and instructions to each patient in their group, and the time and resources required if they would like to use the SAMP test in their practice (see Appendix 19) (Read et al. 1987). Similar to patients, a validated Likert scale with nine response categories (1-9) was used for each question (Avis and Smith 1994).

5.2.11 Sample size

No example of sample size estimation for the evaluation of the practical properties (acceptability and feasibility) of physical performance OMs were identified in the literature. The COSMIN checklist suggests that a small sample size (15-30) is sufficient in this phase of validating a newly developed OM (de Vet et al. 2011). However, the SAMP test is a physical performance test and it is unethical and inappropriate to use it to test one group of participants three times since this could create avoidable distress to them as they are already coping with NS-NP and possible ULD (Henley and Frank 2006). It was also impractical to request the participants to attend three different sessions. Therefore, it was proposed to recruit a larger sample size and to use a stratified sampling procedure to allocate theses participants into three balanced groups, in demographic characteristics and symptoms severity, and SAMP testing each group

using one of the proposed weights only in a single testing session. Consequently, a list of 80 potential participants was obtained from the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital, of which 70 patients were found eligible and willing to participate in the study, and thus included in the analysis.

5.2.12 Data analysis

Data was transferred into SPSS (IBM SPSS Statistical Software, version 24.0) for statistical analysis. In order to meet the objectives of this study, simple descriptive analysis using frequencies, percentages, means, standard deviations (SD), minimum and maximum scores were used to present: demographic information (e.g. occupation, NS-NP duration, weight and height), the severity of the neck and UL symptoms (before testing, immediately after testing and 24 hours after testing), and the NDI scores. The SAMP test scores using the three proposed weights (½-kg, 1-kg, 1½-kg) and the patient's view after SAMP testing were descriptively analysed to: (1) determine the most appropriate weight to be used in the practical application of the SAMP test, which addressed the first objective (see Section 5.1.1), (2) assess the patients' acceptability regarding the instruction and performance of the SAMP test and (3) assess the feasibility (patients' burden) regarding the time and effort required when performing the SAMP test. The examiners' opinions after testing were descriptively analysed to assess their acceptability regarding the length and complexity of the SAMP tasks as well as the feasibility (clinicians' burden) regarding the resources required when using the SAMP test.

5.3 Results

The flow of participants through each stage is presented in Figure 5.2. A list of 80 patients was obtained from the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital. Following the phone screening, 70 out 80 patients were eligible. Eight patients were ineligible and 2 patients declined to participate. Following the face-to-face assessment, all 70 patients were eligible for SAMP testing, happy to participate in the study and consented in writing.

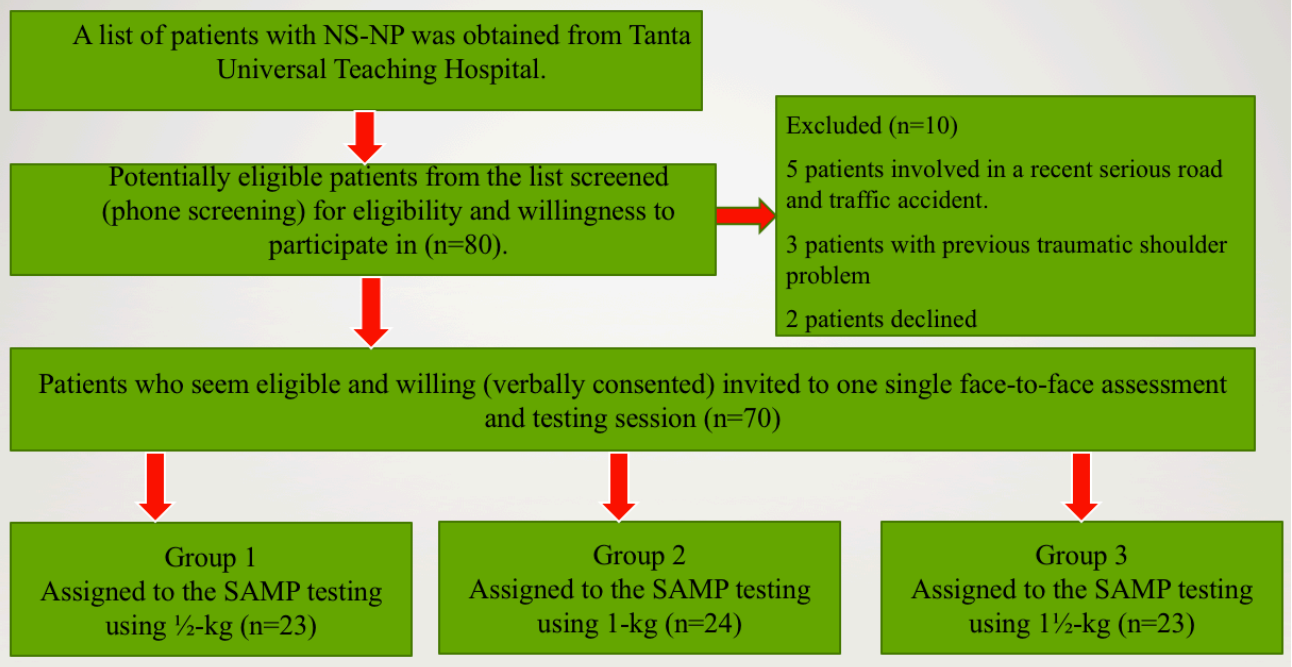


Figure 5.2 Flow-Chart of Participants in Acceptability and Feasibility Study

5.3.1 **Demographic characteristics**

Following the stratification procedure, the 70 participants were randomly allocated to one of three groups for SAMP testing, with each group using a different weight (½-kg, 1-kg, 1½-kg). Demographic and participant characteristics (e.g. age, occupation, weight and height) alongside the duration of NS-NP are presented in Table 5.1. Comparison between the three groups regarding demographic characteristics indicated that the mean age was slightly, but not significantly, higher in group 3 (40.87 years) compared with the mean age in group 1 (39.13 years) and group 2 (39.92 years), but the standard deviation and minimum/maximum of age were almost identical. The sample was balanced across the three testing groups. However, group 1 had a higher proportion of participants with acute NS-NP (13%) compared with those in groups 2 (0%) and 3 (0%). Group 3 had a slightly higher proportion of participants with chronic NS-NP (78.2%) compared with group 1 (56.7%) and group 2 (62.5).

Table 5.1 Demographic characteristics

Variables	Group 1 Tested Using (½kg) N=23	Group 2 Tested Using (1kg) N=24	Group 3 Tested Using (1½kg) N=23
Age (Years)			
Mean	39.13	39.92	40.87
SD	4.576	4.403	4.818
Minimum	34	35	35
Maximum	50	50	50
Occupation: Frequencies (%)			
House-Wife	8 (34.8)	9 (37.5)	10 (43.5)
Office Clerk	14 (60.9)	14 (58.3)	13 (56.5)
Techer	1 (4.3)	1 (4.2)	0 (0)
NS-NP Duration: Frequencies (%)			
0-5 Weeks (acute pain)	3 (13)	0 (0)	0 (0)
6-11 Weeks (sub-acute pain)	7 (30.3)	9 (37.5)	5 (21.8)
12+Weeks	13 (56.7)	15 (62.5)	18 (78.2)
Weight (kg): Frequencies (%)			
74-80	2 (8.7)	0 (0)	0 (0)
81-85	2 (8.7)	2 (8.3)	0 (0)
86-90	7 (30.6)	10 (41.7)	10 (43.5)
91+	12 (52)	12 (50)	13 (56.5)
Height (cm): Frequencies (%)			
155-160	13 (56.5)	14 (58.3)	18 (78.3)
161-165	8 (34.9)	10 (41.7)	5 (21.7)
166+	2 (8.6)	0 (0)	0 (0)

SD: Standard Deviation, NS-NP: Non-Specific Neck Pain.

5.3.2 SAMP scores, symptoms severity, acceptability and feasibility for patients and examiners

All participants in this study completed the NDI questionnaire (Arabic version) on their arrival, and the severity of their neck and upper limb symptoms were measured before testing, immediately after testing and 24 hours after testing. The acceptability and feasibility of the testing procedure were evaluated immediately after SAMP testing for all participants individually. Aggregated test data describing the SAMP scores, NDI scores, the neck and UL symptoms severity alongside patients' acceptability and feasibility regarding the SAMP testing procedure across the three testing groups are presented in Table 5.2.

Participants in group 1 who were tested using the ½-kg had a higher average score 21 reps/30s in SAMP testing, whereas those in group 3 who were tested using 1½-kg recorded lower average scores 10 reps/30s in the SAMP testing. However, participants in group 2 who were tested using 1-kg reported an average score 16 reps/30s in the SAMP testing. The neck and UL symptoms severity increased immediately after testing on the VAS scale (0-10) across the three-testing groups. This was resolved 24 hours after testing for the participants who were tested using the ½-kg weight and 1-kg

weight, whereas those who were tested using the 1½-kg weight were still sore 24 hours after testing. Those who were tested using the 1-kg weight improved slightly with regard to the severity if their UL symptoms 24 hours after testing (before testing: mean=2, immediately after testing: mean=5, 24-hours after testing: mean=1).

Participants in group 1 reported that the ½-kg weight was extremely light or moderately light when used in the SAMP testing, whereas those in group 3 reported that the 1½-kg weight was moderately heavy or neither heavy nor light when used in the SAMP testing. However, those in group 2 reported that the 1-kg weight was slightly light or neither heavy nor light when used in the SAMP testing. Participants across the three-testing group reported that the SAMP testing procedure was extremely easy in relation to instruction and performance, and highly suitable in relation to the time and effort required regardless of the weight used.

After testing, the examiners (n=3) involved in the data collection of this study were requested to provide their qualitative opinions about the SAMP testing procedure (acceptability and feasibility). They all confirmed that the SAMP test was extremely easy or very easy to use in relation to providing an explanation with demonstration, and the overall administration and completion. They also agreed that the SAMP test was highly appropriate regarding the resources required (e.g. time and cost) and that there was no need for additional training, or extra staff or technological support to facilitate the application of the SAMP testing procedure. Patients in group 1 had an average SAMP score of 21 reps/30 seconds and those in group 2 had average SAMP score of 16 reps/30 seconds, whereas those in group 3 had an average SAMP score of 9 reps/30 seconds. The SAMP's administration and completion time for groups 1 and 2 ranged between 50 to 60 seconds (administration) and 110 to 120 seconds (completion) respectively, whereas for group 3 it ranged between 30-60 seconds (administration) and 90-120 seconds (completion).

Descriptive data regarding the SAMP testing administration and completion time; the SAMP scores; NDI scores; and neck/UL symptoms severity before testing, immediately after testing and 24 hours after testing alongside both the participants' and the examiners' opinions about the SAMP testing procedure for the three groups of patients are presented in Appendix 20.

Table 5.2: SAMP scores, symptoms severity and patients' acceptability/feasibility

		Group 1: ½-kg (N=23)	Group 2: 1-kg (N=24)	Group 3: 1½-kg (N=23)
SAMP scores				
	Mean (SD)	21 R (7.064)	16 R (7.433)	10 R (5.579)
	Range	23 R	21 R	20 R
	Min	10 R	6 R	0 R
	Max	33 R	27 R	20 R
NDI Score: Total score=100				
	Mean	35/100	45/100	43/100
	SD	15.254	17.107	15.111
NDI severity categories (frequencies)				
	Low	N=7	N=7	N=6
	Moderate	N=10	N=9	N=9
	Severe	N=5	N=6	N=6
	Extremely severe	N=1	N=2	N=2
Neck symptoms severity: Mean (SD) VAS 0-10				
0=No Symptoms	Before testing	4/10 (0.733)	4/10 (0.721)	4/10 (0.656)
10=Worst Possible Symptoms	Immediately after testing	5/10 (1.041)	5/10 (0.794)	6/10 (0.778)
	24 hours after testing	3/10 (0.885)	3/10 (0.776)	5/10 (0.869)
Upper limb symptoms: Mean (SD)				
VAS 0-10	Before testing	2/10 (0.928)	2/10 (1.056)	2/10 (0.728)
0=No Symptoms	Immediately after testing	4/10 (0.853)	5/10 (1.142)	6/10 (0.765)
10=Worst Possible Symptoms	24 hours after testing	2/10 (0968)	1/10 (1.213)	5/10 (0.984)
Patient acceptability: Weight				
Likert scale 1-9	Range	3/9	5/9	5/9
1=Extremely Light	Min	1/9	2/9	4/9
9=Extremely Heavy	Max	4/9	7/9	9/9
Patients acceptability: Willingness and ability				
Likert Scale 1-9	Range	4/9	5/9	4/9
1=Extremely Easy	Min	1/9	2/9	5/9
9=Extremely Difficult	Max	5/9	7/9	9/9
Patients acceptability: Instruction and performance				
Likert Scale 1-9	Range	1/9	0/9	0/9
1=Extremely Easy	Min	1/9	1/9	1/9
9=Extremely Difficult	Max	2/9	1/9	1/9
Patients burden/feasibility: Time and effort				
Likert Scale 1-9	Rang	2/9	3/9	3/9
1=Highly Suitable	Min	1/9	1/9	1/9
9=Completely Unsuitable	Max	3/9	4/9	4/9

SD: Standard Deviation, SAMP: Single Arm Military Press, R: Repetition, NDI: Neck Disability Index, VAS: Visual Analogue Scale, Min: Minimum, Max: Maximum.

5.4 Discussion

5.4.1 Summary and discussion of the main findings

This chapter reports the findings from a pragmatic randomised controlled study that investigated the practical properties of the SAMP test. This study has captured information about the acceptability and feasibility of the SAMP test for both patients and clinicians. Consequently, the aim and objectives of this validation of the SAMP test have been met (see Section 5.1.1).

Nearly all patients and clinicians involved in this study agreed that regardless of the weight used, the SAMP test hand weight was simple, quick, inexpensive and extremely easy to use in relation to instruction, performance, and time and effort required to administer the test and score performance. It would appear that the SAMP test is an acceptable physical performance test for patients as well as for clinicians. The feasibility of the SAMP test was established regarding the time and resources required. The reasons for this high acceptability and feasibility are that the SAMP test is convenient, since it can be efficiently administered by physiotherapists and/or any other individual of varying experience in any setting using minimum equipment (one dumbbell). Further, it is time effective as it only takes up to 2 minutes for administration and completion. However, the ½-kg hand weight was considered by the majority of patients in group 1 to be too light, hence they had a high SAMP average score. Using a light hand weight in the application of a physical performance test will risk missing out on the identification of patients with subtle/mild pain and disability (Fitzpatrick et al. 1998). This suggests that the ½-kg hand weight is unsuitable for use in female patients with NS-NP. Conversely, the 1½-kg hand weight was considered by the majority of patients in group 3 to be too heavy, and hence they had a low SAMP average score. In addition, some patients were either unwilling or unable to lift the weight and the majority of patients in group 3 had increased neck and UL symptoms immediately after testing, and they were still sore 24 hours after testing. Using a heavy hand weight in the application of a physical performance test will distress patients and risk aggravating pain and disability for a longer period after testing, which could lead to patient fear and avoidance of the intervention and consequently, non-adherence to rehabilitation (Ahuga 2015). The findings suggest that the 1½-kg hand weight is unsuitable for use in female patients with NS-NP. The 1-kg hand weight was considered by the majority of patients in group 2 to be neither too light nor too heavy, thus they had an average score in the SAMP testing. In addition, nearly all patients in group 2

reported an improvement in their neck and UL symptoms severity 24 hours after testing. This indicates that the 1-kg hand weight is suitable for use in female patients with NS-NP. The findings of this study lead to the conclusion that 1-kg SAMP test is acceptable, feasible and therefore suitable for use by female patients with NS-NP.

The findings of this study are consistent with those of other studies which used 1-kg hand weights when examining the functional capacity of the UL for patients with shoulder pathology or NS-NP (MacDermid et al. 2007, Kumta et al. 2012, Constand and MacDermid 2013). In their validation study, MacDermid et al. (2007), developed the Functional Impairment Test-Hand, and Neck/Shoulder/Arm (FIT-HaNSA), a new PBOM for measuring the functional capacity of UL in patients with shoulder pathology. The test protocol consists of 3 subtasks, of which each task can be continued for up to 5 minutes. In the first task, a shelf is placed at the participant's waist level and a second shelf is placed 25 cm above it, while three 1-kg containers/bottles are placed 10 cm apart on the lower shelf. Using the affected arm, the participants are instructed to lift the 3 containers, one at a time, from one shelf to the other. In the second task, the shelves are adjusted so one shelf is placed at the participant's eye level and the second is placed 25 cm below it. Using their affected arm, the patients are instructed again to lift the three 1-kg containers between the shelves. In the final task, a shelf is placed at the participant's eye level with an attachable plate perpendicular to the shelf and projecting out towards the participant. Using their affected arm, participants are instructed to repeatedly screw and unscrew bolts in a specific pattern. The test was developed and different scores was observed when comparing the healthy subjects to either the surgical-list patients with shoulder impingement or a variety of mild shoulder pathology patients, and this indicates that the 1-kg hand weight was suitable when testing the functional capacity of the UL in patients with mild, moderate or severe shoulder pathology.

In their case control study, Constand and MacDermid (2013) recruited 7 patients with NS-NP and 12 healthy subjects to investigate the level of difficulties patients with NS-NP may experience when performing reaching overhead and reading tasks. Participants in this study completed two tasks that incorporated different types of neck movement, reach overhead tasks to represent upper cervical motion and long neck flexion, and reading tasks to represent lower cervical spine motion. In the first task, a shelf was placed 64 cm above the participant's naval and a second shelf was placed slightly

higher than the participant's head. Using their dominant hand, from the standing position, participants were instructed to reach overhead by moving a 1-kg container from the lower shelf and place it on the higher shelf repeatedly for 30 seconds. This indicates that the 1-kg hand weight was suitable when testing the UL functional capacity (overhead reaching) in patients with NS-NP.

5.4.2 Strengths of the study

This study was conducted, analysed and interpreted in accordance with the ISOQOL checklist recommendation for the evaluation of the measurement and practical properties of OMs (Reeve et al. 2013), alongside the new **CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments" (COSMIN) recommendations regarding the evaluation of content validity for health-related OMs (Terwee et al. 2018) in order to ensure a robust methodology. Key stakeholders, patients and clinicians were involved in this study, which is essential when validating OMs to ensure relevance, acceptability and feasibility (Fitzpatrick et al. 1998). The study achieved more than the recommended sample size (15-30) according to the COSMIN checklist (de Vet et al. 2011). This could be because appropriate methods for Egyptian patients (phone screening) were used initially to invite patients to participate in the study. Another reason could be because participants were requested to attend one single assessment and testing session. The implementation of broad inclusion and exclusion criteria and standardised assessment ensured that the included participants were representative of female patients with different types of NS-NP (e.g. acute, subacute and chronic) and the degrees of symptom severity experienced. Reliable, valid and responsive scales, including the VAS scale (0-10) and Likert scale (1-9), were used to assess the neck and UL symptoms severity before testing, immediately after testing, and 24 hours after testing, and to evaluate the acceptability and feasibility of the SAMP testing procedure to both patients and clinicians (Donovan et al. 1993, Fitzpatrick et al. 1998, Hasson and Arnetz 2005, Brokelman and Haverkamp 2012). The use of a stratified randomisation strategy to allocate participants led to three broadly similar testing groups.

5.4.3 Limitations of the study

This study was conducted on female patients and this may prevent the generalisability of the findings to male patients with NS-NP. It is well known that, on average, males are stronger than females but most of the difference in strength is based on body size and muscle cross-sectional area only (Hunter 2010). Consequently, the 1-kg hand

weight may be too light and therefore not suitable for male patients with NS-NP. However, the SAMP test is a performance-based and include tasks to evaluate the strength and endurance of the UL and women tend to have better muscle endurance than men since they generally take longer to fatigue (Cheng et al. 2003). In addition, the findings of other studies which have used a1-kg weight when examining the functional capacity of the UL in patients with shoulder pathology or NS-NP indicate that a 1-kg hand weight was suitable for female as well as male patients with shoulder pathology or NS-NP (MacDermid et al. 2007, Kumta et al. 2012, Constand and MacDermid 2013).

This study focused on participants with NS-NP which limits the generalisability to other forms of NP, such as cervical radiculopathy and Whiplash Associated Disorders (WAD). It is likely that patients with cervical radiculopathy and WAD may have more severe neck problems and greater levels of central sensitisation, and consequently may experience greater levels of ULD. Using the 1-kg SAMP test may be too difficult for these groups. Therefore, the suitability of the 1-kg SAMP test would need to be established in a separate study.

5.4.4 Clinical implications

This study established the acceptability and feasibility of the SAMP test. From the patient perspective, the SAMP test is extremely easy regarding instruction and performance; time and effort required; and physical ability and willingness. Therefore, the SAMP test is suitable for use in practice. From the clinicians' perspective, the SAMP test is advantageous in relation to qualities such as demonstration, instruction, score, interpretation of score and overall administration and completion. No additional training, staff or technological support are required to facilitate the application (administration and completion) of the SAMP test. Incorporating a 1-kg hand weight in the practical application of the SAMP test was most appropriate for female patients with NS-NP regardless of the severity of their symptoms. The test is likely to elevate the neck and UL symptoms severity slightly but they return to normal or possibly reduce slightly after 24 hours, and this indicates that Delayed Onset Muscle Soreness following testing is likely to be minimal. Consequently, the SAMP test can be recommended as a suitable test for use in clinical practice. However, this test has not yet been shown to be a reliable or valid measure of UL capacity in patients with NP, and therefore cannot at this stage be recommended as a measure of UL functional capacity in female patients with NS-NP.

5.4.5 Research implications

Although the findings of this chapter have demonstrated that the SAMP test is a highly acceptable and feasible measure of UL functional capacity in female patients with NS-NP; further research is required to investigate important measurement properties such as reliability and validity. In addition, future testing should be done to establish its feasibility, reliability and validity in male patients and patients with other NP disorders, including cervical radiculopathy and WAD.

5.5 Conclusion

The 1-kg SAMP test was found to be an acceptable and feasible weight for use with female patients with NS-NP and therefore has demonstrated its potential for use in clinical practice. The measurement and practical properties of the SAMP test should be confirmed and further tested in other female populations, male populations and in different types of NP.

The next chapter is a validity study which addresses the final research question (see Section 1.5) and describes the evaluation of the reliability and validity of the SAMP test in a second cohort of Egyptian female patients with NS-NP and healthy subjects.

Chapter 6: Measuring upper limb disability in female patients with non-specific neck pain: Evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test.

6.1 Introduction

This chapter describes the methodology, and presents and discusses the results, of a study that was conducted to establish the reliability and validity of the 1-kg SAMP test in Egyptian female patients with NS-NP and healthy subjects.

Chapter five of this thesis has successfully established the acceptability, feasibility and therefore the suitability of the SAMP test for clinical practice as a measure of UL functional capacity for female patients with NS-NP. However, this chapter also concluded that further testing for reliability and validity would be required before the SAMP test can be formally recommended for use in clinical practice.

6.1.1 Aim

The main aim of this study was to investigate the reliability and validity of the SAMP test as a measure of ULD Egyptian female patients with NS-NP alongside healthy subjects.

6.1.2 Hypotheses (reliability)

1. The inter- and intra-rater reliability of the SAMP test will be high (ICCs ≥ 0.90)
2. The agreement in the repeated measurement will be very high and the SEM will be very low (SEM ≤ 1) and smaller than the smallest detectable change (SDC).

6.1.3 Hypotheses (validity)

1. The SAMP test performance has highly significant ($p < 0.05$) and substantial negative correlation ($r > -0.70$) with the DASH score (convergent validity).
2. The SAMP test has the capacity to discriminate between healthy subjects and a patient group.
3. The SAMP test has the capacity to discriminate between patient groups of different severity of NS-NP.

6.1.4 Reliability and validity: Concepts

6.1.4.1 Reliability

Reliability is an essential requirement of all outcome measures; poor reliability alongside a high level of measurement error would limit the extent to which the findings of an instrument can be generalised. Consequently, this would reduce the usefulness as well as the clinical utility of the instrument (de Vet et al. 2006). Reliability concerns the extent to which the measurement of stable patients can be reproduced when the same instrument is used at different moments, in different conditions, by different examiners or by the same examiner at different times (Streiner and Norman 2003). Reliability as a domain reflects the extent of correlation as well as the agreement in repeated measurements and comprises three measurement properties: internal consistency, reliability and measurement error (Mokkink et al. 2010b). Internal consistency is assessed only for PROMs and is defined as the extent to which items in a questionnaire are interrelated (Mokkink et al. 2010b). It concerns the extent to which items assess the same construct in a unidimensional scale of a multi-item instrument (de Vet et al. 2011). Given that the SAMP test is a physical performance test, internal consistency is not relevant and therefore redundant in this study. Reliability as a measurement property is described as the proportion of the total variance in the measurement resulting in the consistency of the scores as well as the error which is not attributed to true changes but resulting in the systematic and random error of a patient's scores (Mokkink et al. 2010b). It includes test-retest, inter-rater and intra-rater reliability which are illustrated in Table 6.1. Reliability examines the ability of an instrument to distinguish between patients despite the measurement errors that are related to the variability between the study objects "participants" and is expressed in correlations using the Interclass Correlation Coefficient (ICC) (de Vet et al. 2006). The ICC uses a typical basic formula:

$$\text{Reliability} = \frac{\text{variability between study objects}}{\text{variability between study objects} + \text{measurement error}}$$

Table 6.1: Reliability types (de Vet et al. 2011)

Inter-rater Reliability	Examines the variation between multiple examiners (2 examiners or more) who measure the same patients/subjects using the same instrument in the same occasion/session.
Intra-rater Reliability	Examines the variation in repeated measurements by the same examiner on stable patients/subjects using the same instrument under the same condition in different occasions/sessions.
Test-retest Reliability	Examines the variation in repeated measurements on stable patients/subjects under the same condition using the same instrument, but the examiner is neglected/not involved (e.g. self-reported survey instrument).

The reliability parameter ICC ranges between 0.0 to 1.0, where values close to 0.0 indicate poor reliability and ICC values close to 1.0 suggest high reliability (Portney and Watkins 2009). An interpretation of the ICC values is illustrated in Table 6.2.

Table 6.2: ICC values interpretation (Landis and Koch 1977

ICC Value	Interpretation of Strength
< 0.000	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.59	Moderate
0.60-0.79	Substantial (High)
0.80-1.00	Almost Perfect (Very High)

ICC: Interclass Correlation Coefficient

Measurement error is a measurement property of the reliability domain which represents the systematic and random error of a patient's score that is not attributed to true changes in the construct being measured (Mokkink et al. 2010b). Measurement error quantifies the extent to which an OM provides accurate scores, independent from the population and reflects the agreement in repeated measurements (de Vet et al. 2011). It is the absolute measurement error over repeated measurements of the test when the patients are stable between measurements. Measurement error is expressed by the Standard Error of Measurement (SEM), which estimates how the repeated measurement of a patient on the same instrument tends to be distributed around their "true" score. SEM is the standard deviation of the errors of measurement that are associated with an instrument's scores and is equal to the square root of the error variance ($\sqrt{\sigma^2}$ error) (de Vet et al. 2006). Low levels of SEM indicate high levels of score accuracy and high levels of SEM indicate low levels of score accuracy (Vincent and Weir 2012). SEM should be smaller than the Smallest Detectable Change (SDC), which represents the minimal change that a patient must show on an OM to ensure that the observed change is real and not just measurement error (Bland and Altman 1996). The ICC and SEM formulas and variance are illustrated in Table 6.3.

Table 6.3: ICCs, SEM and SDC formulas and variances (de Vet et al., 2011)

ICC	Interclass Correlation Coefficient
SEM	Standard Error of Measurement
SDC	Smallest Detectable Change
σ^2	Variance: The statistical term that is used to represent the variability in the measurement scores
σ^2 p	Variance due to differences in the study objects (participants)
σ^2 pt.	Variance due to systematic differences between examiners/physiotherapists (i.e. pt. 'A' and pt. 'B').
σ^2 error (residual)	Variance due to differences in the interaction between participants and examiners.
ICC agreement	= σ^2 p ÷ σ^2 p + σ^2 pt. + σ^2 residual
ICC consistency	= σ^2 p ÷ σ^2 p + σ^2 residual
SEM agreement	= √ σ^2 pt. + σ^2 residual
SEM consistency	= √ σ^2 residual
SDC	= 1.96 X √2 X SEM

σ^2: Variance, pt.: Physiotherapist

6.1.4.2 Validity

Validity as a domain concerns the degree to which an instrument truly measures the construct for which it was developed and validated to measure and comprises three measurement properties: content, criterion, and construct validity (Mokkink et al. 2010b). Content validity, concerns the adequacy between the content of an instrument and the construct being measured and is not relevant in this study since it is usually an aspect of OMs development (Mokkink et al. 2010b). Criterion validity describes the degree to which the scores of an instrument adequately reflects the scores of a gold standard (that is, a perfectly valid assessment/OM that is considered to represent the true state of the construct being measured) (Mokkink et al. 2010b). Criterion validity is used to provide evidence of an OMs validity only when a gold standard is available (de Vet et al. 2011). Given that there is currently no gold standard for measuring UL functional capacity in patients with NS-NP, criterion validity is redundant in this study. Construct validity is the other measurement property which should be assessed to provide evidence of the validity of an instrument when a gold standard of the construct being measured is not available. It concerns the degree to which the scores of the instrument under study are consistent with clearly and a priori formulated hypotheses regarding the relationship with the scores of other instruments that should be measuring the same construct (Mokkink et al. 2010b). Construct validity is frequently investigated using analyses that test for statistical differences (de Vet et al. 2011). Construct validity types and definitions are presented in Table 6.4, and the Pearson correlation values and interpretation are presented in Table 6.5.

Table 6.4: Construct validity types and definitions (Mokkink et al. 2010b)

Construct Validity	The extent to which the scores of an instrument are consistent with clearly a priori formulated hypotheses.
Convergent Validity	The extent to which an instrument scores are correlated with the scores of other instruments that measure the same construct based on a priori formulated hypotheses
Discriminate Validity	The extent to which an instrument has the capacity to discriminate between groups that are known to be clinically different.

Table 6.5: Pearson correlation values and interpretation (Domholdt 2000)

Pearson value	Interpretation of strength
$r = < 0.40$	Low
$r = 0.40$ to 0.70	Moderate
$r = > 0.70$	High

r: Pearson Correlation Coefficient

6.2 Methods

6.2.1 Study design

A large-scale validation study was designed to investigate the reliability and validity of the SAMP test in female Egyptian patients with NS-NP and healthy subjects in accordance with the "**CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments" (COSMIN) checklist recommendations (Mokkink et a. 2010a, Terwee et al. 2012). Patients participants were recruited from the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital in Egypt, while matched healthy subjects were recruited from the general population living in Egypt. Participants (patients and healthy subjects) were tested using the 1-kg SAMP test since it was found to be acceptable and feasible for use in patients with NS-NP (explored in chapter 5, sections 5.3 and 5.4). This study was conducted from March 2016 to April 2016 in the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital (Egypt).

6.2.2 Ethical approval

This study was approved by Tanta Universal Teaching Hospital (Rheumatology and Physical Therapy Department) (see Appendix 7). Subsequently, approval was sought and gained from the Health & Wellbeing Faculty Ethics Committee at Sheffield Hallam University to conduct this study in Egypt (SHUREC) (see Appendix 8).

6.2.3 Study setting

This study, similar to the acceptability and feasibility study in chapter 5, took place in the Rheumatology and Physical Therapy Department at Tanta Universal Teaching

Hospital in El-Gharbia province, Egypt. Further details about the setting can be found in (chapter 5, section 5.2.3).

6.2.4 Study sample

Female patients who had visited the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital or were on the waiting list with a diagnosis indicating NS-NP were recruited to this study alongside frequency matching of healthy subjects. Participants in this study were different from those who participated in the validation (SAMP acceptability and feasibility) study that is reported in chapter 5.

6.2.5 Inclusion and exclusion criteria

6.2.5.1 Patient participants

The inclusion and exclusion criteria for patient participants in this study were identical to the validation study (SAMP acceptability and feasibility) described in chapter 5 and are described in detail in sections 5.2.5.1 and 5.2.5.2.

6.2.5.2 Healthy subjects

Healthy subjects were considered for inclusion in this study if they were a female adult aged ≥ 18-years, with no history of head/neck/UL trauma and no current or recent neck or UL problems (within the last three months). Eligible participants were frequency matched with prospective patient participants regarding gender, age, occupation, weight and height.

6.2.6 Recruitment of study sample

Similar to the validation study described in chapter 5, a list of 300 female patients with NS-NP was obtained from the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital. Healthy female subjects were recruited from the general population by announcement via social network sites (e.g. Facebook and Twitter), personal networking and posters/flyers within Tanta University and Tanta city centre. Willing and potentially eligible participants were asked to contact the researcher (ASEA) using the research hotline number. This resulted in 100 healthy subjects contacting the study hotline after which the line was closed and the process was stopped.

Potentially eligible patients and healthy subjects were telephoned by the researcher (ASEA) or another member of the research team who explained the study and conducted a phone screening to confirm their provisional eligibility and gain verbal consent. In the phone screening, patients and healthy subjects were checked against the inclusion and exclusion criteria. A telephone checklist of clinical and demographic questions was completed for each patient and healthy subject to ensure standardisation of the information given and to cover the inclusion and exclusion criteria (see Appendix 21). If provisional eligibility was confirmed and verbal consent was obtained, an appointment for a face-to-face assessment followed by the first testing session was organised. The company that was used in the validation study in chapter 5 was commissioned again in this study to deliver the study information sheet (see Appendix 22) to each verbally consented participant at least 48 hours before their first assessment and testing session. Participants were requested to carefully read the study information sheet and discuss their potential participation with their family and friends.

Participants (patients and healthy subjects) who were found provisionally eligible and verbally consented to take part in this study were booked for a face-to-face assessment at Tanta Universal Teaching Hospital (Egypt). Participants were given the opportunity to ask questions, the researcher checked eligibility against the inclusion/exclusion criteria prior to gaining written consent from willing participants. Participants were informed that taking part in this study was entirely voluntary and that they could withdraw from the study at any time without reprisal.

6.2.7 Sample size

Sample size estimates in this study were based on the COSMIN checklist recommendations, which suggest that at least 50 patients are required for a reliability study in order to achieve a reasonable number of dots on the Bland and Altman plot which estimates the limits of agreement in the repeated measurements (de Vet et al., 2011). COSMIN also suggests that a larger sample size (≥100) is better when evaluating the inter-rater and intra-rater reliability of an instrument in order to obtain a Confidence Interval (CI) > 0.90 around Interclass Correlation Coefficient (ICC) of 0.90 – 0.95 (Giraudeau and Mary 2001). However, given that the validity (convergent and discriminant) of the SAMP test are statistically investigated in this study using a group of healthy subjects and four sub-groups of patients with different severity level of NS-

NP, it was proposed to recruit ≥ 200 patients and 50-100 healthy subjects (de Vet et al., 2011).

6.2.8 Clinical staff involved in the SAMP testing

A total of four female examiners who were all physicians and employed at the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital, with at least 3-years of experience in working with musculoskeletal patients were involved in the recruitment as well as the data collection in this study.

6.2.9 Training and information delivered to staff

The three examiners involved in the data collection of the validation study described in chapter 5, plus one additional examiner (four examiners in total) were involved in the data collection in this study. Twenty-four hours prior to the face-to-face assessment and the first testing session, the four examiners attended a 45-minute practical training and information giving session at the Rheumatology and Physical Therapy Department at Tanta Universal Teaching Hospital (Egypt). The session was delivered by the PhD researcher (ASEA) who discussed the purpose of this reliability and validity study, provided a brief outline of the SAMP test description and the practical application; demonstrated the warm-up; standard utilisation of the SAMP technique; and recognition of compensatory strategies and/or ineligible lift, and the scoring system. Examiners were paired for simultaneous SAMP testing of patient participants (inter-rater reliability). The first pair, a member of staff (GE) was the rater who administered the SAMP test, while a second member of staff (HA) was the co-assessor who only recorded the SAMP test score independently but simultaneously in the first testing session. The second pair, a member of staff (BH) was the rater, while the fourth member of staff (AH) was the co-assessor in the first testing session. The role of the rater and co-assessor were shifted in the second testing session for the two pairs to examine whether a switch between the rater and co-assessor (interaction with patients during the administration) could present another variance and influence the performance of patients (de Vet et al. 2006). Data regarding the SAMP testing were collected and analysed for the two pairs of examiners.

6.2.10 Outcome measures

The SAMP test and two PROMs (NDI and DASH) were used in this study. A brief description of each instrument used are given below.

6.2.10.1 The SAMP tests

The SAMP test performance consists of tasks that simulate daily activities of carrying, lifting and using the UL in overhead function. The attribute of interest in this instrument related to the sustained work that involves repetitive overhead activity. The SAMP test procedure was conducted as described in chapter 5 (see Section 5.2.7). First, participants were asked to stand with their feet positioned at shoulder width and lift a 1-kg dumbbell, using their dominant hand, to shoulder level (see Figure 6.1A). Second, participants were asked to raise their hand with the dumbbell directly overhead by extending through the elbow (see Figure 6.1B). Finally, participants were asked to repeat this process as fast and as frequently as possible for 30 seconds but to take their pain and fatigue into account. It was emphasised that they could stop at any time during the 30 seconds testing. The test was stopped if a participant reported extreme pain or fatigue. The scoring system of the SAMP test is a repetition count (number of repetitions) within 30 seconds, in which higher values represent batter performance and a lower level of ULD.

Fig 6.1A *Fig 6.1B*

Figure 6.1 SAMP Test Protocol

6.2.10.2 The Neck and Disability Index (NDI)

The NDI is a standard PROM for measuring a patient's disability due to neck pain and is the most commonly used instrument in clinical practice and research (MacDermid et al. 2009, MacDermid et al. 2013). The NDI has 10 items, in which 7 items are related to activities of daily living, 2 items related to pain, and 1 item related to concentration

(Vernon and Mior 1991). Each item is scored from 0-5 and the total score is expressed as a percentage score, with higher scores indicating greater disability. The NDI is supported by the largest volume of neck related clinical measurement evidence and demonstrates excellent measurement properties across multiple studies (MacDermid et al. 2009). In addition, the NDI is the most commonly used patient-reported outcome measure for patients with neck pain and/or disability in clinical practice and research practice (Linton 2000, Sterling et al. 2003, Dunckley et al. 2005, Abrams et al. 2006, Bot et al. 2007, Cote et al. 2008, de Koning et al. 2008, Nordin et al. 2008, Skeat and Perry 2008). The NDI was translated and culturally-adapted to the Arabic language and its reliability and validity were determined in Arabic-speaking patients with NP (Shaheen et al. 2013). The NDI scoring intervals for interpretation, which was used to create the 4-subgroups of patients in this study, are shown in Table 6.6.

Table 6.6: NDI score and interpretation (Vernon and Mior 1991

NDI Score	Interpretation
0-8	No Disability
10-28	Mild Disability
30-48	Moderate Disability
50-68	Severe Disability
69-100	Extremely severe Disability

6.2.10.3 The Disability of the Arm, Shoulder and Hand (DASH)

The DASH is a multidimensional PROM developed primarily to evaluate the upper limb disability and/or symptoms as a single functional unit (Hudak et al. 1996). The DASH uses 30-items related to difficulty when performing activities which use the upper limb. The dimension physical function comprised 21-items, pain 5-items and emotional/social function comprised 4-items. Each item is scored on a 1-5 scale. A total score is calculated by summing item scores and transforming them into a score from 0-100 where 0 equals no disability and 100 equals the most severe disability (Hudak et al. 1996). Since its development, the measurement properties of the DASH questionnaire have been extensively and successfully evaluated for a variety of upper limb conditions and translated and cross-culturally adapted into over 40 different languages, including Arabic (Hudak et al. 1996, Turchin et al. 1998, Beaton et al. 2001, Westphal et al. 2002, Veehof et al. 2002, Soohoo et al. 2002, Offenbacher et al. 2003, Greenslade et al. 2004, Liang et al. 2004, Raven et al. 2008). The DASH was also validated to measure ULD in patients with NS-NP (Huisstede et al. 2009, Mehta et al. 2010). Further details about the

DASH and its validation for patients with NS-NP can be found in chapter 2 (see Section 2.3.4).

6.2.11 Testing procedure

Participants who were found to be provisionally eligible and verbally consented to take part in this study were booked for a face-to-face assessment and SAMP testing at Tanta Universal Teaching Hospital (Egypt). The testing procedure, including phone screening is detailed in Table 6.7. In the first assessment and testing session, participants were requested to complete the NDI and the DASH questionnaires (Arabic versions) on their arrival (see Appendices 11and 24), since they were used to confirm a patient's eligibility and facilitate allocation to subgroups. This was followed by a face-to-face assessment with the researcher (ASEA) in a designated room where their weight and height as well as their neck and upper limb symptoms severity were measured alongside subjective examination using standardised clinical questions (see Appendix 25). Participants were then requested to meet immediately with the relevant examiners, as appropriate, for the first SAMP testing, if they were found eligible and consented in writing (see Appendix 26). The SAMP testing was conducted by two examiners independently but simultaneously for each patient participant, meanwhile healthy subjects were tested by one examiner only. Each examiner completed the data collection sheet (see Appendix 27) for each participant regarding the SAMP score (i.e. the number of valid SAMP repetitions within 30 seconds) and the administration time (that is, the time taken for the examiner to describe and demonstrate the SAMP test as well as instruction and performance of the SAMP test by the patient). Two pairs of examiners "GE and HA" and "BH and AH" were used in this study. The first session, including the face-to-face assessment and testing procedure, took up to 60 minutes, after which participants were booked for the second session within 7 days (minimum of 4 days) after the first session as appropriate and convenient to participants.

Participants in this study were stable in the time interval between sessions. During the second session, participants were requested to complete the NDI and the DASH questionnaires (Arabic versions) upon their arrival. The neck and upper limb symptoms severity were measured as per the first session, using the VAS scale 0-10 where 0 indicates no pain/symptoms and 10 indicates the worst possible pain/symptoms. Participants were requested to meet immediately with the relevant examiners for the second SAMP testing. Patients were SAMP tested in the second session by the same

pair of examiners who tested them in the first session, though they had swapped their rater/co-assessor roles. Similar to the first session, healthy subjects were tested by one examiner only that had tested them in the first session. The second session took up to 45 minutes, after which participants were discharged.

Table 6.7: Testing procedure

	What	When
Telephone Call (Preliminary Assessment)	Explanation of the study protocol Phone screening Verbal consent Provision of information sheet	Immediately after the patients were identified by obtaining the patients' list or after the healthy subjects were identified by contacting the research hotline number.
Face-to Face Assessment and First Testing Session	Outcome measures (NDI and DASH) Face-to-Face Assessment Written consent SAMP testing	Arranged to suit participant (patient or healthy subject)
Second Testing Session	Repeat outcome measures (NDI and DASH) N/UP symptoms severity (patient only) Repeat SAMP testing	Within 7 days (minimum 4 days) after the first session.

NDI: Neck and Disability Index, DASH: Disability of Arm, Shoulder and Hand, SAMP: Single Arm Military Press Test, N: Neck, UL: Upper Limb.

6.2.12 Data analysis

Data were transferred into Excel and then to SPSS (IBM SPSS Statistical Software, version 24.0) for further analysis. Descriptive statistics (mean, standard deviation, standard error of mean, and 95% confidence interval) were computed for the SAMP test, the DASH and the NDI for patient participants and healthy subjects.

Inter-rater reliability, which compares the scores of the two independent but simultaneous examiners, was calculated for the two pairs of examiners (see Table 6.8) across the two testing sessions. Intra-rater reliability, which compares the score of a single examiner across two sessions was calculated for the two examiners (GE and BH) across the two testing sessions. Given that the objective was to examine the reliability of the SAMP test and the type of data are continuous, the ICC 2,1 (Modal: Two-Way Random, and Type: Absolute Agreement, and Single Measure) value with 95% Confidence Interval (CI) is recommended for calculating the reliability parameters as it considers both systematic and random errors (Shavelson 1991, de Vet et al. 2011). For inter- and intra-rater reliability, the ICC was anticipated to be (≥ 0.90) (see hypothesis 1 in section 6.1.2). Measurement error was calculated by estimating the SEM which was

derived using a two-way analysis of variance (ANOVA) ICC 2,1 (McGraw and Wong 1996). The SDC was then calculated using the formula: (SDC = 1.96 X √2 X SEM). For measurement error, the SEM was anticipated to be (≤ 1) and smaller than the SDC (see hypothesis 2 in section 6.1.2).

Table 6.8: Examiners and testing sessions

Examiners	Session 1	Session 2
First Pair: Examiner GE	Rater (Administrating the SAMP test)	Co-assessor (counting SAMP score only)
First Pair: Examiner HA	Co-assessor	Rater
Second Pair: Examiner BH	Rater	Co-assessor
Second Pair: Examiner AH	Co-assessor	Rater

GH: Dr Ghada Eid, HA: Dr Heba Abdo, BH: Dr Basma Hassan, AH: Dr Asmaa Hamdy

Since there is no gold standard OM available in relation to measuring UL functional capacity in patients with NS-NP, construct validity (convergent and discriminant) was assessed in this study to provide evidence for the validity of the SAMP test as a measure of UL functional capacity in patients with NS-NP. The construct validity (convergent) for the SAMP test was assessed in terms of the level of association between the SAMP test scores and standardised instrument UL PROM (DASH) that measures the same constructs (physical function). Convergent validity of the SAMP test was elucidated by use of Pearson correlation coefficient (*r*) (de Vet et al. 2011). Significant and high negative correlation between the SAMP score and the DASH score was anticipated (see hypotheses 1 in section 6.1.3).

Discriminant validity (known group) was evaluated by assessing the difference in the SAMP scores between patient participants and healthy subjects. The differences in the SAMP scores for the patient group and healthy subject group was calculated using the independent sample t-test. It was anticipated that the patient group would have a statistically significant ($p < 0.05$) and substantially poorer SAMP score compared to the healthy subject group with a large magnitude of difference in the mean (Effect Size = > 0.8) (Cohen 1988).

Discriminant validity between patient sub-groups was also examined by comparing SAMP test scores obtained during the first session by examiner GE using the

independent sample t-test and an analysis of variance. Four sub-groups of patients were formulated based on the NDI scores as illustrated in Table 6.9. It was anticipated that patients in the extremely severe NS-NP sub-group would have a statistically significant ($p < 0.05$) and substantially poorer SAMP score as compared to the patients in the other 3 sub-groups with mild, moderate or severe NS-NP with a large magnitude of difference in the mean (Effect Size = > 0.8) (Cohen 1988). The effect size was calculated in this study in accordance with Cohen (1988) recommendations using the formula: (Effect Size = $t^2 \div t^2 + (n1 + n2 - 2)$, in which t donated to t-test score, n 1 donated to sample 1 and n 2 donated to sample 2. The effect was considered small when the effect size ranged between 0.0 – 0.4, medium 0.5 – 0.7 and large 0.8+ (Cohen 1988).

Table 6.9: Patient sub-groups

Sub-Groups	Description
Sub-Group 1	Patients with mild NS-NP who scored between 10-29 in the NDI
Sub-Group 2	Patients with moderate NS-NP who scored between 30-49 in the NDI
Sub-Group 3	Patients with severe NS-NP who scores between 50-68 in the NDI
Sub-Group 4	Patients with extremely severe NS-NP who scored between 69-100 in the NDI

NDI: Neck Disability Index

6.3 Results

The flow of patient participants and healthy subjects through each stage is presented in Figure 6.2 below. A list of 300 patients was obtained from the Rheumatology and physical Therapy Department at Tanta Universal Teaching Hospital. Following the phone screening, 250 patients were eligible and willing to voluntarily participate in the study. Thirty patients were ineligible, 20 patients declined and 40 patients did not turn-up for their first assessment and testing session. Following the face-to-face assessment in the first session, 210 patient participants and 81 healthy subjects were found eligible for SAMP testing, interested to participate in the study, consented in writing and participated in session 1 testing. All the participants from session 1 testing participated in session 2 testing (no drop-out).

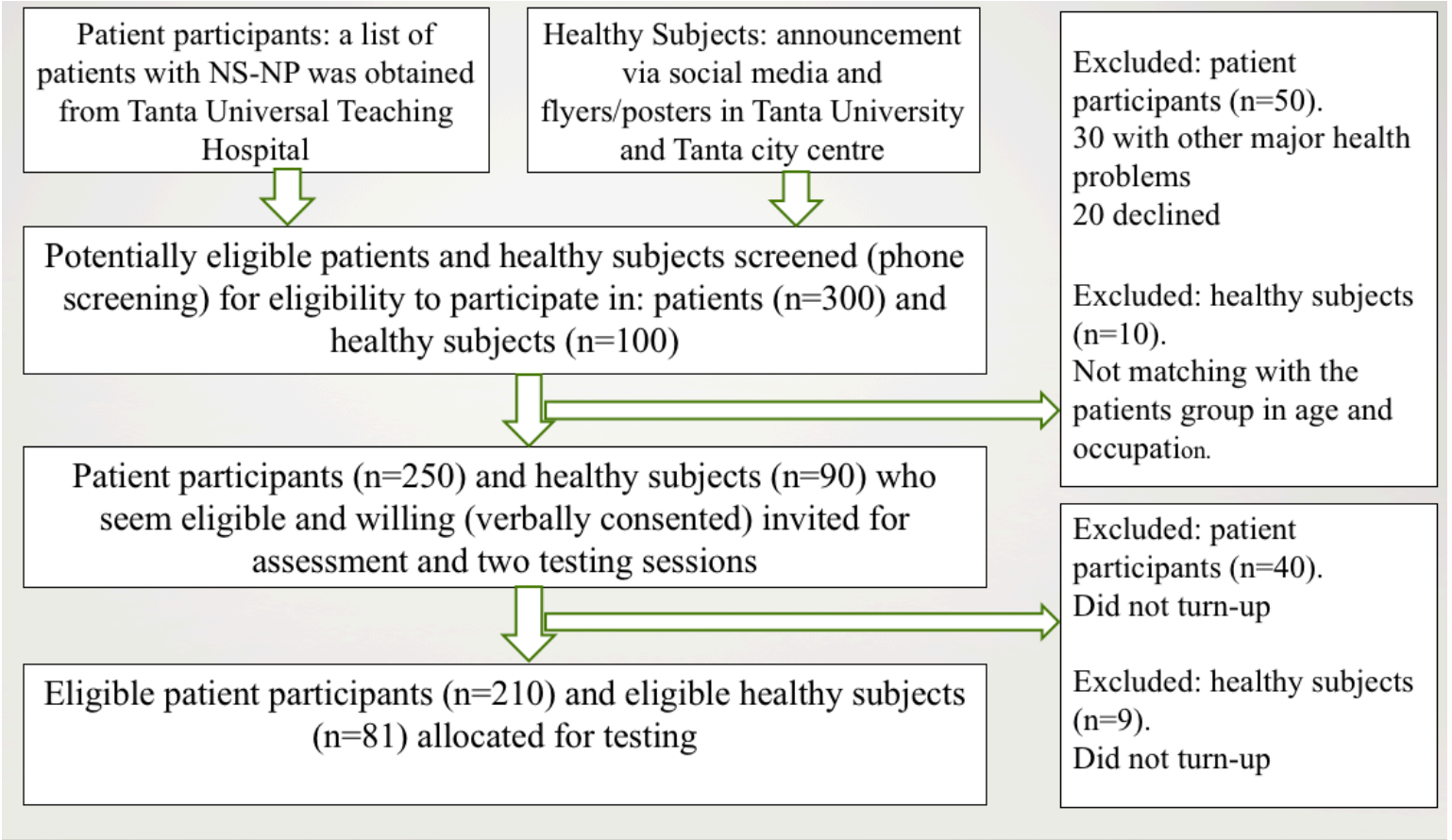


Figure 6.2: Flow Chart of Patient Participants and Healthy Subjects in Reliability and Validity Study

6.3.1 Participants characteristics and baseline data

The demographic characteristics of participants in this study alongside the baseline data are summarised in Table 6.10. The mean age of the recruited 210 patient participants was 40.41 ± 4.938 years. Further, 81 healthy subjects were recruited in the control group (36.54 ± 4.917 years). There were no significant differences between patients and healthy subjects on age, occupation, weight or height, and this indicates that these groups were well frequency matching in demographics. However, as expected, there were clear and substantial differences between these groups regarding the severity of NS-NP and UL functional capacity in all measures. In the second testing session, patients reported slight, but non-significant, improvements on their neck symptoms severity (NSS) and UL symptoms severity (ULSS) scores, indicating that these groups were stable between testing sessions.

Table 6.10: Participants characteristics at baseline stratified by the NDI

Variables	Healthy Subjects N=81	All Patients with NS-NP N=210	Patients with Mild NS-NP: N=23	Patients with Moderate NS-NP: N=120	Patients with Severe NS-NP: N=46	Patient with E Severe NS-NP: N=21
Age in years						
Mean	36.54	40.41	34.43	38.50	44.72	48.43
SD	4.917	4.938	2.609	2.834	3.053	2.336
Minimum	30	30	30	32	41	41
Maximum	50	53	39	46	53	52
Occupation: frequencies (%)						
Office Clark	77 (95.1)	200 (95.2)	17 (73.9)	118 (98.3)	45 (97.8)	20 (95.2)
Teacher	3 (3.7)	8 (3.8)	6 (26.1)	1 (0.8)	0 (0.0)	1 (4.8)
House Wife	1 (1.2)	2 (1.0)	0 (0.0)	1 (0.8)	1 (2.2)	0 (0.0)
Weight (kg): Frequencies (%)						
75-80	4 (5)	10 (4.8)	1 (4.4)	5 (4.2)	3 (6.5)	1 (4.8)
81-85	4 (5)	11 (5)	1 (4.4)	6 (5)	2 (4.4)	2 (9.6)
86-90	23 (28.3)	60 (28.6)	7 (30.4)	33 (27.5)	10 (21.8)	9 (42.8)
91+	50 (61.7)	129 (61.4)	14 (60.8)	76 (63.3)	31 (67.3)	9 (42.8)
Height (cm): Frequencies (%)						
155-160	25 (30.8)	65 (31)	8 (34.7)	36 (30)	15 (32.6)	6 (28.6)
161-165	51 (63)	130 (62)	14 (60.9)	76 (63.4)	27 (58.6)	13 (61.8)
166+	5 (6.2)	15 (7)	1 (4.4)	8 (6.6)	4 (8.8)	2 (9.6)
NSS: Sessions 1 and 2						
Mean		4.40 – 4.04	2.57 – 2.43	3.74 – 3.36	5.78 – 5.50	7.19 – 6.52
SD	0	1.475–1.452	0.507 - 0.507	0.642 - 0.754	0.593 - 0.658	0.402 - 0.512
Minimum		2 - 2	2 - 2	3 - 2	5 – 4	7 - 6
Maximum		8 - 7	3 – 3	5 - 5	7 - 7	8 – 7
ULSS: Sessions 1 and 2						
Mean		2.45 – 2.28	1.00-1.00	1.73 - 1.61	3.78 – 3.70	5.19 – 4.43
SD	0	1.414-1.295	0.000 - 0.000	0.645 - 0.677	0.593 - 0.465	0.402 - 0.507
Minimum		1 - 1	1 - 1	1 – 1	3 – 3	5 – 4
Maximum		6 - 5	1 – 1	3 - 3	5 – 4	6 - 5
NDI Scores: (Session1)						
Mean	4.63	43.38	25	36.85	55.98	73.19
SD	0.798	14.474	3.275	5.291	5.053	3.326
Minimum	4	20	20	30	50	69
Maximum	6	80	29	49	67	80

NSS: Neck Symptoms Severity, ULSS: Upper Limb Symptoms Severity, SD: Standard Deviation, NDI: Neck Disability Index, E: Extremely, NS-NP: Non-Specific Neck Pain

6.3.2 Descriptive statistics of the SAMP test

Participants in the study were SAMP tested in two sessions approximately 1 week apart. Descriptive statistics of the SAMP test scores for the healthy subject group, all patient groups, and the four sub-groups of patients are summarised in Table 6.11. There were significant differences between the healthy subject group and the all patients group, of which the healthy subjects group scored substantially higher in the SAMP performance in the two testing sessions. In addition, the sub-group of patients with extremely severe NS-NP demonstrated the poorest SAMP performance across the two testing sessions.

Table 6.11: Descriptive statistics of the SAMP test

	Healthy Subjects N=81	All Patient NS-NP N=210	Mild NS-NP N=23	Moderate NS-NP N=120	Severe NS-NP N=46	Extremely Severe NS-NP N=21
Session 1						
Mean	35.23	17.90	24.17	20.48	13.15	6.67
SD	3.348	6.167	2.588	4.048	2.996	1.713
Std. Error of Mean	0.372	0.426	0.540	0.370	0.442	0.374
Minimum	28	3	19	12	7	3
Maximum	39	30	28	30	20	9
Session 2						
Mean	35.07	17.99	24.04	20.60	13.17	7.00
SD	2.692	6.140	2.549	4.123	2.961	1.673
Std. Error of Mean	0.299	0.424	0.532	0.376	0.437	0.365
Maximum	29	3	20	12	7	3
Maximum	40	30	28	30	20	9

NS-NP: Non-Specific Neck Pain, SD: Standard Deviation, Std. Error of Mean: Standard Error of Mean.

6.3.3 Descriptive statistics of the DASH and the NDI questionnaires

Participants in this study completed the NDI and the DASH questionnaire before SAMP testing in each session. The DASH and the NDI scores for all participants are presented in Table 6.12. The DASH and the NDI scores for the all patients group were significantly higher compared to those in the healthy subject group in the two testing sessions. Furthermore, the DASH and NDI scores across the two sessions for the patient sub-group with extremely severe NS-NP were also significantly higher compared to those in the sub-groups with mild, moderate or severe NS-NP, indicating the strong relationship between NS-NP and ULD.

Table 6.12: Descriptive statistics of the DASH and NDI questionnaires

Variable	Healthy Subjects N=81	All Patient N=210	Mild NS-NP N=23	Moderate NS-NP N=120	Severe NS-NP N=46	Extremely Severe NS-NP N=21
DASH: Session 1						
Mean	4.04	31.66	16.87	23.11	45.54	66.33
SD	1.167	16.420	1.632	7.346	5.648	5.083
Std. Error of Mean	0.130	1.133	.340	0.671	0.833	1.109
Minimum	3	15	15	15	40	57
Maximum	6	75	20	47	62	75
DASH: Session 2						
Mean	4.10	30.56	16.74	22.22	43.98	64.00
SD	0.682	15.810	1.514	6.988	5.467	5.099
Std. Error of Mean	0.076	1.091	0.316	0.638	0.806	1.113
Minimum	3	14	15	14	38	55
Maximum	6	74	19	45	60	74
NDI: Session 1						
Mean	4.63	43.38	25.00	36.85	55.98	73.19
SD	0.798	14.474	3.275	5.291	5.053	3.326
Std. Error of Mean	0.089	0.999	0.683	0.483	0.745	0.726
Minimum	4	20	20	30	50	69
Maximum	6	80	29	49	67	80
NDI: Session 2						
Mean	4.48	38.61	21.57	31.48	51.67	69.38
SD	0.550	14.934	2.501	5.689	6.332	3.930
Std. Error of Mean	0.061	1.031	0.522	0.519	0.934	0.858
Minimum	4	18	18	22	42	64
Maximum	6	78	26	45	65	78

NS-NP: Non-Specific Neck Pain, SD: Standard Deviation, Std. Error of Mean: Standard Error of Mean.

6.3.4 Inter-and intra-rater reliability

The ICC, SEM and SDC with 95% Confidence Interval (CI) for the patient group were calculated for the two pairs of examiners across the two testing sessions to assess the inter-rater reliability and agreement for the SAMP scores for all patient group and the four sub-groups of patients. Meanwhile, the ICC, SEM and SDC with 95% CI, patient group, were calculated for examiners A (GE) and B (BH) across the two sessions to assess the intra-rater reliability and agreement for the SAMP scores for all patient group and the four sub-groups of patients. The ICC 2,1, SEM and SDC statistics with 95% CI (lower bound and upper bound) for inter-and intra-rater reliability are presented in Tables 6.13 and 6.14 respectively. The ICCs exceeded 0.90 for all patients group and the four sub-groups with NS-NP (see hypothesis 1, section 6.1.2) and therefore the SAMP test demonstrated almost perfect reliability (Landis and Koch 1977). The SEM was ≤ 1 and smaller than the SDC for inter- and intra-rater reliability (see hypothesis 2, section 6.1.2) indicating that the SAMP performance demonstrated high levels of score accuracy and agreement (de Vet et al. 2011, Vincent and Weir 2012).

Table 6.13: Inter-rater Reliability Coefficient, SEM and SDC with 95% CI for the SAMP test

Variables	All Patient with NS-NP N=210	Patients with Mild NS-NP N=23	Patients with Moderate NS-NP: N=120	Patients with Severe NS-NP: N=46	Patients with E Severe NS-NP: N=21
Session: (1)					
ICC 2,1	0.995	0.951	0.983	0.999	0.999
95% CI (LB – UB)	0.993 - 0.996	0.884 - 0.980	0.972 - .990	0.998 - .999	0.998 - .999
SEM	0.42	0.54	0.48	0.10	0.10
SDC	1.2	1.5	1.3	0.28	0.28
Session: (2)					
ICC 2,1	0.997	0.950	0.992	0.999	0.983
95% CI (LB – UB)	0.996 - 0.998	0.888 - 0.978	0.998 - 0.994	0.998 - 0.999	0.958 - 0.993
SEM	0.35	0.58	0.37	0.10	0.21
SDC	1.0	1.6	1.0	0.28	0.58

ICC: Interclass Correlation Coefficient, NS-NP: Non-Specific Neck Pain, E: Extremely, CI: Confidence Interval, LB: Lower Bound, UB: Upper Bound

Table 6.14: Intra-rater Reliability Coefficient, SEM and SDC with 95% CI for the SAMP test

Variables	All Patient with NS-NP: N=210	Patients with Mild NS-NP N=23	Patients with Moderate NS-NP: N=120	Patients with Severe NS-NP: N=46	Patients with E Severe NS-NP: N=21
Examiner: A					
ICC 2,1	0.997	0.964	0.992	0.999	0.926
95% CI (LB – UB)	0.996 - 0.998	0.918 - 0.984	0.989 - 0.995	0.998 - 0.999	0.784 - 0.972
SEM	0.35	0.49	0.35	0.10	0.41
SDC	1.0	1.2	1.0	0.28	1.1
Examiner: B					
ICC 2,1	0.994	0.938	0.983	0.999	0.893
95% CI (LB – UB)	0.998 - 0.996	0.858 - 0.973	.0 956 - 0.991	0.998 – 0.999	0.675 – 0.960
SEM	0.44	0.62	0.45	0.10	0.48
SDC	1.2	1.7	1.2	0.28	1.3

ICC: Interclass Correlation Coefficient, NS-NP: Non-Specific Neck Pain, E: Extremely, CI: Confidence Interval, LB: Lower Bound, UB: Upper Bound.

6.3.5 Construct validity (convergent)

The SAMP test score for examiner (A) in the first testing session was used in assessing construct validity. To test the a priori formulated hypotheses regarding the convergent validity of the SAMP test, the relationship between the SAMP test and other extensively validated and commonly used OMs that measure the same construct (physical function) for the UL, namely DASH, was investigated using Pearson correlation (*r*). Pearson correlation and the *p* values are summarised in Table 6.15. Highly significant negative correlations, exceeding -0.70 were observed between the SAMP performance and the DASH scores in the expected direction. This indicates that the SAMP test and the DASH are closely related instruments and measure the same construct, which support the hypothesis (see hypothesis 1 in section 6.1.3) and established the SAMP test's convergent validity (Domholdt 2000).

Table 6.15: Correlation between the scores on the SAMP and the PROM (DASH)

Measurements	Healthy Subjects: N=81	Patients with NS-NP: N=210
DASH v SAMP		
r	-0.870	-0.911
p	< 0.001	< 0.001

DASH: Disability of the Arm, Shoulder and Hand, SAMP: Single Arm Military Press, r: Pearson Correlation, p: p value.

6.3.6 Construct validity (discriminant)

The SAMP test scores for examiner (A) in the first testing session was used in assessing discriminant validity. To test the a priori formulated hypotheses regarding discriminant validity of the SAMP test, an independent t-test was conducted to compare the SAMP scores between the healthy subject group and the all patient group; the patient sub-groups one (mild NS-NP) and two (moderate NS-NP); subgroups two and three (severe NS-NP); and sub-groups three and four (extremely severe NS-NP). Meanwhile, an analysis of variance (ANOVA) was used to compare the SAMP scores between the four patient sub-groups. The mean, standard deviation, mean difference, t-test (t), ANOVA (F), effect size and *p* value with 95% confidence interval (lower bound – upper bound) between the groups are presented in Table 6.16. The healthy subjects had statistically significant higher mean SAMP score than those subjects with NS-NP ($p < 0.001$) with a large magnitude of difference in the mean, exceeding 0.8 in the anticipated direction. This support the hypothesis (see hypothesis 2, section 6.1.2) and indicates that the SAMP test has the capability to discriminate between healthy subjects and patients with NS-NP. Patients with extremely severe NS-NP had statistically significant lower SAMP scores than those with severe, moderate or mild NS-NP ($p < 0.001$) with a large magnitude of difference in the mean, exceeding 0.8 in the anticipated direction. This support the hypothesis (see hypothesis 3, section 6.1.3) and indicates that the SAMP test has the capacity to discriminate between patient sub-groups with different symptoms severity levels.

Table 6.16: Differences in the SAMP scores for patients/healthy subjects and patient sub-groups

Group	Mean (SD)	Mean Difference	*t*	*f*	*p*	ES	95% CI (LB – UB)
Patients (N=210) Controls (N=81)	17.90 (6.167) 35.23 (3.348)	-17.339	-23.964		< 0.001	0.67	(-18.763 – -15.915)
Mild (N=23) Moderate (N=120)	24.17 (2.588) 20.48 (4.048)	3.699	4.214	20.857	< 0.001	0.11	(2.382 – 5.016)
Moderate (N=120) Severe (N=46)	20.48 (4.048) 13.15 (2.996)	7.323	12.715		< 0.001	0.50	(6.181 – 8.464)
Severe (N=46) E. Severe (N=21)	13.15 (2.996) 6.67 (1.713)	6.486	11.208		< 0.001	0.66	(5.329 – 7.642)

SD: Standard Deviation, t: Independent Sample T-Test, f: ANOVA, p: p value, CI: Confidence Interval, LB: Lower Bound, UB: Upper Bound, ES: Effect Size, E: Extremely.

6.4 Discussion

6.4.1 Summary and discussion of the main findings

This chapter reports the findings of a validation study and has captured information regarding the reliability (inter- and intra-rater), agreement (measurement error) and construct validity (convergent and discriminant). The findings demonstrate that the SAMP test is a reliable and valid measure of UL functional capacity for female patients with NS-NP, hence meeting the aim set out in section 6.1.1.

6.4.1.1 Reliability

The SAMP test in this study demonstrated almost perfect levels of reliability. Interpretation of the 95% Confidence Intervals around the Interclass correlation coefficient ($ICC_{2,1}$) values suggest that the 'true' estimate of inter-rater and intra-rater reliability of the SAMP test ranges between ICCs of 0.993 and 0.996, indicating a very high degree of stability of the SAMP scores over time and agreement between examiners. This exceeds the ICC ≥0.90 set out in hypothesis 1 of (section 6.1.2). Given that an ICC of at least 0.70 is considered to be satisfactory for an instrument to detect differences in severity between groups in research practice and an ICC value of 0.90-0.95 is required to enable this instrument to detect differences in severity between individual patients in clinical practice (de Vet et al. 2011). This indicates that the SAMP test can be consistently well used by different examiners or the same examiner in different occasions to measure the UL functional capacity in female patients with NS-NP in clinical practice and research practice. The reliability results of this study have confirmed previous results reported for the SAMP test. McLean et al. (2010a) investigated the reliability (inter- and intra-rater) of the SAMP test in a series of preliminary studies on a symptomatic and asymptomatic non-patient population

(n=265) and reported an ICC of 0.94 and 0.99, indicating almost perfect reliability. The reason for this high level of reliability may be due to the simplicity, efficiency and standardisation of the SAMP test procedure. The SAMP test requires simple instructions and minimal training for observers who are required to only count the valid repetitions within 30 second in order to complete the administration of the test.

The SAMP test also demonstrated very low levels of measurement error. The Standard Error of Measurement (SEM) was very low, smaller than the Smallest Detectable Change (SDC), as expected and ranged between 0.10 and 0.58 for inter-rater reliability and 0.10 and 0.62 for intra-rater reliability, indicating a very high level of precision in the patients' scores. This is smaller than the SEM ≤ 1 and smaller than the SDC set out in hypothesis 2 of (section 6.1.2). In conclusion, these results support the hypotheses of this study in terms of reliability and suggests that the SAMP test can be considered a reliable instrument for use in clinical practice as well as research practice to evaluate the UL functional capacity in female patients with NS-NP. The hypotheses regarding reliability and agreement have been confirmed (see Section 6.1.2).

6.4.1.2 Validity

Construct validity was used in this study to determine the validity of the SAMP test since there is no gold standard available in relation to measuring UL functional capacity in patients with NS-NP. The SAMP test demonstrated very high level of convergent validity. The Pearson correlation analysis revealed a significant ($p < 0.001$) and substantial negative correlation ($r = 0.911$) between the SAMP score and the DASH score in the patient group. The correlation and significance levels were in the anticipated direction stated in hypothesis 3 (section 6.1.3), indicating that the SAMP test and the DASH are measuring a related construct and thus providing evidence of convergent validity for the SAMP test. The results of this study in terms of convergent validity confirmed the previous results reported for the SAMP test. McLean et al. (2010a) investigated convergent validity of the 3-kg SAMP test in a series of preliminary studies on a symptomatic and asymptomatic non-patient population and reported a highly significant negative correlation between the SAMP test scores and the DASH scores ($r = 0.814$, $p < 0.001$, n=190), indicating a high level of convergent validity.

The SAMP test also demonstrated very high level of discriminant validity. The independent t-test analysis revealed substantial differences in scores between patient participants (mean SAMP score = approximately 18 repetitions) and healthy subjects (mean SAMP score = approximately 35 repetitions). The magnitude of the difference in the mean (mean difference = -17.339, 95% CI: -18.763 to -15.915) with large effect (Effect Size = 0.67), indicates that the SAMP test can consistently distinguish well between patients with NS-NP and healthy subjects (see hypothesis 3, section 6.1.3). In addition, substantial significant differences were observed between the four sub-groups of patients, depending on the severity of their NS-NP (mild, moderate, severe and extremely severe) as the mean SAMP scores were approximately 24, 20, 13 and 7 repetitions respectively. This indicates that the SAMP test can consistently discriminate well between groups of patients with different severity levels (see hypothesis 4, section 6.1.3). In conclusion, these results support the hypotheses of this study in terms of construct validity (convergent and discriminant) and suggest that the SAMP test is a valid measure of UL functional capacity for female patients with NS-NP. The hypotheses about convergent and discriminant validity which are set out in section 6.1.3 have been confirmed.

6.4.2 Strengths of the study

This study was conducted, analysed and interpreted in accordance with the COSMIN recommendations for developing health-related OMs (Mokkink et al. 2010a, Terwee et al. 2012). Independent but simultaneous examiners were used when assessing the inter-rater reliability in order to reduce or possibly prevent the risk of fatigue or soreness to patients, which could lead to drop-out and also to avoid the Hawthorne effect (de Vet et al. 2011). The large sample size achieved (n=290), which was significantly higher than the recommended sample size by the COSMIN checklist (n=100), increased the statistical power of the test of mean differences, prevented potential masking of systematic error and enabled appropriate quantification of the SAMP test reliability and agreement (de Vet et al. 2006). The use of broad inclusion and exclusion criteria and standardised assessments which ensured that the included participants were representative of typical healthy subjects from the general population and patients with a variety of NS-NP severity levels. All patients and healthy subjects who attended the first assessment and testing session were retrained for the second testing session (no drop-out), The reasons for this may be due to the strategies used during the recruitment and data collection stages to ensure participation adherence to the testing protocol, such

as establishing tracking system to locate participants, creating a welcoming environment that made the assessment and testing as smooth and enjoyable as possible, educating patients regarding the importance and the benefits of this testing which may have enhanced their ability and encouraged their adherence, maintaining flexibility when scheduling appointments and using reminder phone calls about appointments. To ensure robust methodology, the DASH questionnaire, which is relevant, standardised and extensively validated UL PROM, was used as comparator when evaluating the convergent validity of the SAMP test.

6.4.3 Limitations of the study

The limitations of this study are very similar to those of the acceptability and feasibility study outlined in chapter 5 and relate to poor generalisability to the male population as well as those patients with other types of neck disorders. This study also involved female participants in the age group (30-50-year) only, which may limit generalisability of the findings to other younger and older patients age groups. In addition, a small number of raters were used when investigating the inter-and intra-rater reliability, which may limit the generalisability of the findings. These limitations point to the requirement for further validation studies in these populations and additional raters.

6.4.4 Clinical implications

The findings of this study provide preliminary evidence that the SAMP test is a reliable and valid measure of UL functional capacity in a female patient with NS-NP. The study also provides normative data for the SAMP performance in healthy subjects which can provide a possible target for rehabilitation. This study also demonstrated that there are significant and substantial differences in the SAMP performance depending on the NS-NP severity level. Healthy subjects had an average SAMP performance of 36, patients with mild NS-NP had an average of 25, patients with moderate NS-NP had an average of 20, patients with severe NS-NP had an average of 14, whereas those with extremely severe NS-NP had an average of 7 repetitions within the 30 seconds. This indicates that the SAMP performance is poorer when the severity increases. Clinicians can use the SAMP test in clinical and research practice to evaluate the UL functional capacity in female patients with NS-NP.

6.4.5 Research implications

In this study, the SAMP test has been subject to only reliability and validity testing in female patients with NS-NP but provides preliminary evidence that the 1-kg SAMP test is a potentially suitable measure for use in clinical practice and research. Further research is required to investigate the measurement and practical properties of the SAMP test in younger and older female patients, male populations and those with other neck disorders such as Whiplash Associated Disorder (WAD), cervical radiculopathy and post-surgical neck disorders. In addition, longitudinal studies to explore the responsiveness of the SAMP test are also warranted.

6.5 Conclusion

This validation study established that the SAMP test has adequate reliability and agreement levels in a female patient population with NS-NP to be used in clinical practice and research practice. The study provides preliminary evidence regarding the expected relationships and convergent validity of the SAMP test with selected standardised instruments measuring the same construct (DASH). The SAMP test was able to discriminate between patients and healthy subjects as well as between NS-NP patient sub-groups with different levels of NS-NP severity.

The next chapter is the concluding chapter to the thesis and aims to bring together the findings of the thesis to show that the overall aim and objectives of the PhD programme have been met and to make recommendations regarding the assessment and management of ULD in patient populations with NS-NP.

Chapter 7: Summary, discussion and conclusion

7.1 Introduction

This chapter summarises and discusses the key findings of the thesis, provides reflection on the main strengths and limitations of the programme of research and explores the implications for clinical practice and future research, before drawing final conclusions.

This thesis was concerned with the clinical measurement and management of ULD in female patient populations with NS-NP. Although ULD is known to be present in people with NS-NP, little was known about the extent to which physiotherapists might measure or manage ULD in this patient group. However, one of the key challenges to optimising the management of this group of patients was the lack of guidance around available measures of ULD which were suitable for use in this population. Consequently, the current programme of research was designed and conducted in order to better understand the challenges of measuring and managing ULD in patients with NS-NP and provide guidance and possible solutions to overcome the challenges.

The aims of this programme of research were to:

1. Investigate the measurement and practical properties of all available ULD OMs that have been developed or validated for patients with NS-NP and identify those that are suitable for use in research and clinical practice.
2. Investigate the measurement and management strategies used by UK physiotherapists for patients with NS-NP, particularly those related to measuring and managing ULD.
3. Identify an acceptable and feasible SAMP test weight for use in female patents with NS-NP.
4. Investigate the measurement properties of the SAMP test in female patients with NS-NP.

7.2 Summary of the thesis methodology

The findings and conclusions reached in this thesis were reported through four research studies. First, a systematic review identified, summarised and critically examined all available studies on the measurement and practical properties of OMs that had been developed or validated to measure ULD in patients with NS-NP (Aim 1). Second, a

literature review investigated current evidence-based management practices recommended for the management of patients with NS-NP. The findings from this review were used to inform the development of the subsequent UK national survey of neck pain, which established current UK physiotherapy about the measurement and management of NS-NP and ULD (Aim 2). Third, a validation study explored the acceptability and feasibility of the SAMP test in female patients with NS-NP (Aim 3). Finally, a validation study investigated the reliability and validity of the SAMP test in female patients with NS-NP and healthy subjects (Aim 4). A summary and discussion of the key findings are presented in the next two sections.

7.3 Summary of the thesis findings

In chapter two, the systematic review identified five clearly defined and reproducible OMs which were supported by 11 developmental and/or evaluative studies. Evidence for the five identified and reviewed OMs in this systematic review was either limited, unknown or unavailable, and this prevented a clear recommendation for any of the identified instruments. However, since brevity of an OM is essential for busy clinical practice, the QuickDASH, NULI and the SAMP test were considered promising ULD OMs for patients with NS-NP, if adequately validated.

Chapter three reported the findings from a literature review that explored current evidence-based management practices within the scope of physiotherapy for patients with NS-NP. A wide range of treatment approaches are currently recommended for the management of patients with NS-NP. Evidence for the effectiveness of these approaches were mostly limited, inconclusive or does not exist when used in isolation. However, evolving evidence suggests the benefits and clinical usefulness of incorporating a multimodal approach to management. Strong evidence of effectiveness was only found for the multimodal management approach that includes exercise and manual therapy with/without patient education programme (Gross et al. 2007, Hurwitz et al. 2008, Kay et al, 2012). The findings of this literature review were subsequently used to inform the development of the UK national survey of neck pain (see chapter 4).

Chapter four reported the findings from the UK national survey that explored musculoskeletal physiotherapists' use of treatment approaches and OMs in the management of patients with NS-NP. The primary findings relating to utilisation of OMs revealed that over one-third of the survey respondents did not utilise any OMs in

the management of their patients with NS-NP. The most commonly reported reasons for this were a lack of clear guidance about the suitability of the available OMs and a lack of time. Further, of the two-thirds of the survey respondents who reported utilising OMs, the majority were consistently using single-dimensional numeric pain and range of motion rating scales (e.g. the Visual Analogue Scale and the Goniometric Measure of Neck Motion). Pain and range of motion do not adequately reflect the construct of NP and therefore indicates an inadequate level of measurement activity within physiotherapists who manage patients with NS-NP. Physical and functional limitations, psychological distress and reduced quality of life constructs that are relevant and frequently associated with NS-NP were rarely, if ever, measured. Moreover, the majority of the physiotherapists in this national survey reported that they would not consider using ULD rehabilitation strategies while managing their patients with NS-NP. This survey suggests that physiotherapists in the UK have a long distance to go regarding implementing evidence-based practice when measuring their patients with NS-NP as well as measuring and managing ULD in patients with NS-NP. These findings were consistent with the findings of comparative surveys (Jette et al. 2009, MacDermid et al. 2013).

Chapter five reported the findings from a pragmatic randomised controlled study that explored the acceptability and feasibility of the SAMP test from both the patients and clinicians' perspective. Following comparison of the ½-kg, 1-kg, 1½-kg weights, the 1kg SAMP test was identified, from both the patient and clinician perspective, as being the most suitable weight for use with female patients with NS-NP. The feasibility of the SAMP test was also demonstrated regarding the time and resources required. This study established that the 1-kg SAMP test is an acceptable and feasible measure of ULD for female patients with NS-NP.

Chapter six reported the findings from the validation study that further investigated the reliability and validity of the SAMP test. The study revealed that the SAMP test had adequate levels of inter-rater and intra-rater reliability, and very low error, which indicates high level of agreement and score accuracy. A high level of correlation between the SAMP test and DASH scores of the patient population confirmed the convergent validity of the 1-kg SAMP test. Substantial significant differences between the SAMP scores of healthy controls compared to the population with NS-NP, coupled with further analysis which demonstrated substantial significant differences between the

SAMP scores of four clinically known groups (mild, moderate, severe and extremely severe NS-NP groups) confirmed the discriminant validity of the SAMP test. All a priori formulated hypotheses (see sections 6.1.2 and 6.1.3) regarding convergent validity and discriminant validity were confirmed. This study established that the 1-kg SAMP test is a reliable and valid measure of ULD for female patients with NS-NP.

7.4 Discussion of key findings

This thesis identified the 1kg SAMP test as a reliable and valid performance-based instrument for measuring ULD in patients with NS-NP. The findings of the systematic review in chapter two highlighted that the QuickDASH, NULI and the SAMP test were promising UL measures for patients with NP. The QuickDASH and NULI are PROMs which have the advantages that they are short and can be completed quickly at the clinic or from home; and they enable patients to report their own pain and functional ability alongside the effects of pain and disability on their psychological, psychosocial and quality of life constructs (Bellamy et al., 1997; Lee et al., 2000; Reneman et al., 2002). However, the main disadvantage is that they are subjective and likely to be biased based on a patient's sex, age, race and perception of pain and/or functional limitations (McDowell, 2006). The sources of bias frequently related to the patients over-or under estimation of their physical and functional ability (Rose et al., 2008; Terwee et al., 2006; Stratford et al., 2006; Stratford et al., 2010). Evidence from the literature suggests that patients often have difficulties estimating their ability to perform activities that they did not undertake during the last week because of pain and/or disability, and that their estimates of their performance often exceeded their actual ability (Youn et al. 1996). Other patients perform their daily activities using compensatory mechanisms and this is likely to influence the magnitude of their disability and thus the patient-reported OM score (Heaton and Bamford, 2001; Bialocerkowski, 2002).

The SAMP test was the only identified performance-based OM that was developed specifically to identify and quantify ULD in the assessment of patients with NP and to monitor its progress during rehabilitation (McLean et al. 2010a). Since it is a physical performance test that requires the patient to use multiple joints to physically perform a task that represents some construct of function including endurance and strength, it has a greater likelihood of accurately capturing the presence of any level of disability (e.g. subtle, mild, moderate, severe or extremely severe) (Curb et al. 2006, Pinheiro et al. 2016). The SAMP test is also advantageous because it can be efficiently administered

by clinicians of varying experience, in any setting, using minimal equipment and within less than 2-minutes. The SAMP test is convenient, efficient and inexpensive and therefore has the characteristics to be very useful in clinical practice and research. The only disadvantage is that the SAMP test does not capture information about patients' psychological, psychosocial or quality of life constructs. However, this information can be acquired by using any standardised generic PROM alongside the SAMP test. Therefore, the SAMP test was taken forward for further testing (see Chapters 5 and 6).

The findings from the UK national survey of neck pain highlighted gaps between evidence-based practice and current UK physiotherapy practice regarding the utilisation of OMs in the assessment and during the management of patients with NS-NP. A lack of clear guidance regarding the suitability of available OMs and a lack of time were found to be the main barriers to utilisation. Routine utilisation of standardised OMs is considered to be a fundamental part of physiotherapy rehabilitation and frequently advocated by clinical guidelines and professional bodies as the optimal way for implementing evidence-based practice (Hammond 2000, Rudd et al. 2000, CSP 2005, College of Occupational Therapists 2007, American Occupational Therapy Association 2010). Further, UK organisations such as the HCPC, CSP and the NHS explicitly recommend the routine utilisation of
standardised OMs wherever practicable (NHS, 2010; CSP, 2012; HCPC, 2013). Meanwhile, in standard 12 of the Standards of Proficiency for Physiotherapists (HCPC, 2013), the HCPC suggests that physiotherapists must be able to collect and document qualitative and quantitative data regarding their patient's condition by using standardised OMs. This is to assure the quality of clinical practice by meeting the patient's needs and changes in health, demonstrating the significance of physiotherapy by enabling physiotherapists to prove their impact and cost-effectiveness (HCPC, 2013).

The findings of the survey also highlighted the gap between evidence-based practice and current UK physiotherapy practice in relation to measuring and rehabilitating ULD in patients with NS-NP. There is strong evidence that patients with NS-NP frequently reported ULD and the presence of NS-NP may be a risk factor for the development and progression of ULD (Walker-Bone et al. 2004, Frank et al. 2005, Bot et al. 2005, Huisstede et al. 2006, Rasmussen et al. 2008, Feleus et al. 2008). Further, ULD may lead to NS-NP becoming recurrent, persistent or disabling (Eriksen et al. 1999, Bot et

al. 2005, McLean et al. 2011). Consequently, routine utilisation of suitable ULD OMs in the assessment and during the management of patients with NS-NP is essential since it enables clinicians to quantify the presence of any ULD and include ULD rehabilitation in the management plan, if indicated (McLean et al. 2011, Osborn and Jull 2013).

The findings of the validation studies (chapters 5 and 6) established the acceptability, feasibility, reliability and validity of the 1-kg SAMP test which is quick, easy, inexpensive and efficient. The SAMP test provides both clinicians and patients alike a quick, easy and intuitive way to understand the extent of ULD and the direction of travel towards incorporating ULD rehabilitation as part of the individualised patient-centred approach to management. The SAMP test has the advantage that it may identify a deficit in UL capacity before a patient is aware themselves that they have a deficit and prevent the progression of ULD by undertaking early UL rehabilitation. This will reduce or possibly eliminate the potential of ULD to have a detrimental cyclical effect on the neck and UL, which may then contribute to chronic, persistent NS-NP and ULD (McLean et al. 2010b). The SAMP test, however still has some limitations as an OM since it has been tested only on female patients with NS-NP in the 30-50-year age group using low number of raters. In addition, the SAMP test has not been tested for responsiveness.

7.5 Contribution to knowledge

This thesis has made a significant contribution to the field of ULD in patients with NS-NP. First, through successfully identifying and critically examining all available OMs, this programme has identified a lack of high-quality UL OMs that are suitable for use in patients with NS-NP. However, promising measures were identified that would be clinically useful if supported by further high-quality validation studies, these include the QuickDASH, NULI and the SAMP test (see section 2.4).

Second, the national survey provided empirical evidence regarding the relatively poor utilisation of multimodal measurement and management approaches in patients with NS-NP. This indicates that many physiotherapists are not adhering to high quality research recommendations or evidence-based guidelines for the measurement and management of patients with NS-NP. Also, there is a relative absence of management, and a lack of measurement of ULD by physiotherapists for patients with NS-NP.

Third, the results of the survey have also provided an extensive and innovative online methodology that incorporated popular Social Networking Sites (e.g. Twitter, Facebook, and LinkedIn) and could be used in future web-based surveys. This online methodology, which does not need any organisational permissions, can provide instant access to a national and/or international population of interest and achieve a large sample size at a very low cost.

Fourth, this programme of research has led to the validation of the 1-kg SAMP test as a valid and reliable measure of ULD for patients with NS-NP. It is also acceptable, feasible and therefore suitable for use in clinical practice and research. It is a quick, easy, intuitive measure which requires minimal equipment, training or resources and may therefore have cross-cultural validity.

7.6 Strengths of the programme of research

The main strength of this thesis is related to the investigation of clinical measurement and management of ULD in patients with NS-NP using robust methodological approaches. A wide range of high-quality strategies were used to address the research aim and objectives including a systematic review in chapter 2, a national survey in chapter 4, a validation (acceptability and feasibility) study in chapter 5 and another validation (reliability and validity) study in chapter 6. Theses research strategies were designed and conducted according to standardised and established guidelines and recommendations such as COSMIN, PRISMA, Web-Based Survey Design Standard and ISOQOL. This programme of research as a whole has been carefully developed to ensure that each of the stages conducted supported the development and conduct of the subsequent stages, which helps to ensure rigour of each of the research studies and therefore the overall validity of the findings at each stage. The data collection phase of the UK national survey of neck pain (chapter 4) used a novel, extensive and innovative online methodology incorporating social networking sites such as Twitter, Facebook and LinkedIn to facilitate the recruitment of a large sample size at a very low cost without permission for access.

7.7 Limitations of the programme of research

The specific limitations to each part of the PhD project have been discussed within individual chapters of the thesis. Overall, the primary limitations of the programme of

research relate to generalisability of the findings due to the use of sample of Egyptian female patients with NS-NP who are in the 30-50-year age group alongside the low number of raters in the validation studies of the SAMP test in chapters 5 and 6. However, there were many good reasons for starting with this population. First, NS-NP is the most common form of NP with approximately 80% of patients experiencing this form (Binder 2007, Jull et al. 2008). Second, given that there are differences between the strength of men and women, it was not feasible to use a mixed population, and so acceptability testing and validation testing need to be conducted in either male or female populations. Third, the most recent and well-designed epidemiological studies found that the incidence of NP is higher among females with an increased risk of developing NP until the 35-49-year age group (Freburger et al. 2005, Cote et al. 2008, Sahin et al. 2008, Hog-Jonson et al. 2008, Hoy et al. 2010, Hoy et al. 2014). Fourth, the Egyptian population was chosen because it was easy to access by the lead researcher (ASEA) and cost-effective. Nevertheless, the transferability of the SAMP test to men and patients with potentially more severe forms of NP such as Whiplash Associated Disorder (WAD) or cervical radiculopathy is not possible. Although the SAMP test has a lot of characteristics which suggest that it has cross-cultural validity, the extent of transferability of the 1-kg SAMP test to other female populations needs to be assessed.

Given the constraints of this programme of research, limited investigation into the measurement properties was conducted, which included acceptability and feasibility testing as well as reliability and validity testing using specific age group and low number of raters. The responsiveness which is validity over time was not investigated in this thesis, and therefore the SAMP test's use as a measure of treatment outcome is not yet established.

7.8 Clinical implications

The key emergent findings of this programme of research is that the 1-kg SAMP test is an acceptable and feasible measure of ULD in female patients with NS-NP. The findings from chapter five of this thesis provided evidence for the practical properties, acceptability and feasibility of the SAMP test when using the 1-kg weight in its practical application, and this in turn established its relevance and clinical utility. In addition, the study in chapter six provided preliminary evidence for the reliability and validity of the SAMP test, and therefore the SAMP test can be recommended for use in clinical practice and research to measure ULD in female patients with NS-NP. The

SAMP test is advantageous because it is easy, simple, quick, inexpensive and can be efficiently administered by physiotherapists or clinicians with varying experience in any setting with the minimum of equipment (one hand-weight) in less than 2 minutes.

The majority (80%) of patients with NS-NP report UL functional limitations. Routine utilisation of a suitable UL OM in the assessment is necessary to quantify the presence of any ULD and provide a rational and a target for UL rehabilitation in the management plan (McLean et al. 2011, Osborn and Jull 2013). However, in this thesis (chapter 4), the UK physiotherapists reported inadequate utilisation of OMs in the assessment of their patients with NS-NP. In particular, physiotherapists rarely, if ever, used an UL OM. This insufficient evaluation of patients may contribute to overall inadequate management as more than half of physiotherapists failed to utilise recommended multimodal management approach, and nearly all UK physiotherapists reported not including UL rehabilitation in the management of their patients with NS-NP. They reported a lack of clear guidance regarding the availability of suitable measures and a lack of time to be the main barriers to their utilisation of OMs. However, this programme of work provided the 1-kg SAMP test that can be used quickly and easily by physiotherapists to partially support the management of patients with NS-NP. However, the SAMP test, as a PBOM, does not have some of the advantages which PROMs possess such as measuring patient's psychological, psychosocial and quality of life constructs and to overcome this, it is possible that the SAMP test can be used to measure the physical/function construct of the UL in patients with NS-NP alongside a standardised generic PROM to measure those other relevant constructs.

7.9 Research implications

The findings from this thesis have provided evidence of the acceptability, feasibility, reliability and validity of the SAMP test, which has the capacity to accurately measure the ULD in patients with NS-NP. However, further research into the validation of the SAMP test is warranted. This includes investigating the measurement and practical properties of the SAMP test in younger and older female patients with NS-NP, male populations and those with other neck disorders such as Whiplash Associated Disorder (WAD), cervical radiculopathy and post-surgical neck disorders. In addition, longitudinal studies to explore the responsiveness of the SAMP test are also warranted.

7.10 Conclusion

This programme of research used a variety of research methodologies to identify and critically examine all available UL OMs for patients with NS-NP. Subsequently, this research described substantial gaps in current evidence-based practice of UK physiotherapists regarding the measurement of patients with NS-NP and associated deficits in the measurement and management of ULD in this patient population. Additional research led to the development of a valid and reliable 1-kg SAMP test which is a measure of UL functional capacity for female patients with NS-NP. The 1-kg SAMP test is acceptable, feasible and therefore suitable for use in clinical practice and research. Furthermore, it is quick to administer (less than two minutes), easy to use, interpret, and can be used in any situation where resources are limited. Further research regarding the validation of the SAMP test is still required to investigate its measurement and practical properties in other populations with NP. Following the findings of this thesis, the SAMP test may go some way towards facilitating improved measurement and management by physiotherapists of female patients with NS-NP.

References

ABRAMS, D., DAVIDSON, M., HARRICK, J., HARCOURT P., ZYLINSKI, M., CLANCY, J. (2006). Monitoring the change: current trends in outcome measure usage in physiotherapy. *Man Ther*, **11** (1), 46-53.

AHMED, O., CLAYDON L., RIBEIRO D., ARUMUGAM, A., HIGGS, C., BAXTER, G. (2012). Social media for physiotherapy clinics: considerations in creating a Facebook page. *Physical Therapy Review*, 7316.3d (DOI 10.1179/1743288X12Y.0000000039).

AHUGA, D. (2015). *Determinants of rehabilitation adherence in outpatient musculoskeletal physiotherapy: A mixed methods project.* PhD, Sheffield Hallam University.

ALAHMAD, G., AL-JUMAH, M., DIERICKX, K. (2012). Review of national research ethics regulations and guidelines in Middle Eastern Arab countries. *BMC Medical Ethics,* **13**:34.

ALRENI A., HARROP, D., LOWE, A., POTIA, T., KILNER, K., MCLEAN, S. (2017). Measures of upper limb function for people with neck pain. A systematic review of measurement and practical properties. *Musculoskeletal Science and Practice*, **29**, 155-163.

ALTMAN, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman and Hall.

AMERICAN OCCUPATIONAL THERAPY ASSOCIATION. (2010). Standards of practice for occupational therapy. *Am J Occup Ther*, **64**, S106-111.

ANGST, F., GOLDHAHN, J., et al. (2009). How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the disabilities of the shoulder, arm and hand questionnaire in the strata of symptoms and function and specific joint conditions. *Quality of Life Research*, **18** (8), 1043-1051.

ANGST, F., SCHWYZER, H., et al. (2011). Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis care & research*, **63** (S11), S174-S188.

AARONSON, N. K. (1989). Quality of life assessment in clinical trials: methodologic issues. *Control Clin Trials*, **10**, 195S–208S.

AARONSON, N. K. (1991). Methodologic issues in assessing the quality of life of cancer patients. *Cancer*, **67**, 844–50.

AARONSON, N. K. (1992). Assessing the quality of life of patients in cancer clinical trials: common problems and common-sense solutions. *Eur J Cancer*, **28**A, 1304–7.

AARONSON, N. K., AHMEDZAI, S., BERGMAN, B., BULLINGER, M., CULL, A., DUEZ, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*, **85**, 365–76.

AVIS, N. E. and SMITH, K. W. (1994). Conceptual and methodological issues in selecting and developing quality of life measures. In: Advances in medical sociology (Fitzpatrick R, editor). *London, JAI Press Inc*, 255–80.

BARUCH, Y. (1999). Response rate in academic studies: A comparative analysis. *Human Relations*, **52**, 421–438.

BEATON, D. E., KATZ, J. N., FOSSEL, A. H., et al. (2001). Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther*, **14**, 128–46.

BEATON, D. E., WRIGHT, J. G., et al. (2005). Development of the QuickDASH: comparison of three item-reduction approaches. *The Journal of Bone & Joint Surgery*, **87** (5), 1038-1046.

BEATON, D., BOMBARDIER, C., et al. (2007). Recommendation for the cross-cultural adaptation of the DASH and QuickDASH outcome measures. *Institute for Work & Health*, **1** (1), 1-45.

BEARMAN, D., SHAFARMAN, S. (1999). The Feldenkrais method in the treatment of chronic pain: a study of efficacy and cost effectiveness. *American Journal of Pain Management AJPM*, **1**, 1-7

BELLAMY, N., KIRWAN, J., BOERS, M., BROOKS, P., STRAND, V., TUGWELL, V., et al. (1997). Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol*, **24**, 799–802.

BERNHARDSSON, S., OBERG, B., JOHANSSON, K., NILESN, P., LARSSON, M. (2015). Clinical practice in line with evidence? A survey among primary care physiotherapists in western Sweden. *Journal of Evaluation in Clinical Practice*, **21**, 1169-1177.

BERTOZZI, L., GARDENGHI, I., TURONI, F., VILLAFANE, J., CAPRA, F., GUCCIONE, A., et al. (2013). Effect of therapeutic exercise on pain and disability in the management of chronic nonspecific neck pain: systematic review and meta-analysis of randomised trials. *Physical Therapy*, **93**, 1026-1036.

BERNAARDS, C. M., ARIENS, G. A., et al. (2008). Improving work style behavior in computer workers with neck and upper limb symptoms. *J Occup Rehabil,* **18** (1), 87-101.

BLAND, J. M. and ALTMAN, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **1**, 307–10.

BLAND, J. M. and ALTMAN, D. G. (1966). Measurement Error. *British Medical Journal*, **313**, 744-753.

BIALOCERKOWSKI, A. (2002). Difficulties associated with wrist disorders-a qualitative study. *Clin Rehab*, **16**, 429–440.

BINDER, A. (2007). The diagnosis and treatment of nonspecific neck pain and whiplash. *Eura Medicophys*, **43** (1), 79-89.

BINDMAN, A. B., KEANE, D., LURIE, N. (1990). Measuring health changes among severely ill patients. The floor phenomenon. *Med Care*, **28**, 1142–52.

BHAMBHANI, Y., ESMAIL, S., BRINTNELL, S. (1993). The Baltimore Therapeutic Equipment Work Simulator: Biomechanical and Physiological Norms for three attachments in healthy men. *The American Journal of Occupational Therapy*, 19-25.

BOND, A. and LADER, M. (1974). The use of analogue scales in rating subjective feelings. *British Journal of Medical Psychology*, **47**, 211-218.

BOLTON, J. E. and WILKINSON, R. C. (1998). Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. *J Manipulative Physiol Ther*, **21** (1), 1-7.

BORGHOUTS, J., KOES, B., et al. (1998). The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain*, **77** (1), 1-13.

BORGHOUTS, J. A., KOES, B. W., VONDELING, H., BOUTER, L., M. (1999). Cost-of-illness of neck pain in The Netherlands in 1996. *Pain*, **80** (3), 629-36.

BOSNJAK, M., TUTEN, T. L., and WITTMANN, W. W. (2005). Unit (non)response in Web-based access panel surveys: An extended planned-behavior approach. *Psychology and Marketing*, **22**, 489–505.

BOUCHET, C., GUILLEMIN, F., BRIANÇON, S. (1996). Nonspecific effects in longitudinal studies: impact on quality of life measures. *J Clin Epidemiol*, **49**, 15–20.

BOWLING, A. (1995). *Measuring disease*. Buckingham: Open University Press.

BOT, S., van der WAAL, J., et al. (2005). Predictors of outcome in neck and shoulder symptoms: a cohort study in general practice. *Spine,* **30**, 495-470.

BOT, S. D., TERWEE, C. B., et al. (2007). Work-related physical and psychosocial risk factors for sick leave in patients with neck or upper extremity complaints. *Int Arch Occup Environ Health*, **80** (8), 733-741.

BROKELMAN, R. B. G., HAVERKAMP, D., van LOON, C., HOL, A., et al. (2012). The validation of the visual analogue scale for patient satisfaction after total hip arthroplasty. *Eur Orthop Traumatol*, **3**, 101-105.

CASPERSEN, C. J., POWELL, K. E., CHRISTENSON, G. M. (1985). Physical activity, exercise, and physical fitness: definition and distinctions for health-related research. *Public Health Reports*, **100** (2), 126-131.

CARLESSO, L., MACDERMID, J., GROSS, A., WALTON, D., SANTAGUIDA, P. L. (2014). Treatment preferences amongst physical therapists and chiropractors for the management of neck pain: results of an international survey. *Chiropractic & Manual Therapies*, **22** (1), 11.

CHILDS, J. D., CLELAND, J. A., ELLIOTT, J. M., et al. (2008). Neck pain: clinical practice guidelines linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *The Journal of Orthopaedic and Sports Physical Therapy*, **38** (9), A1–A34.

CLARE, H., ADAMS, R., MAHER, C. (2004). A systematic review of efficacy of McKenzie therapy for spinal pain. *Aust J Physiother*, **50**, 209-216.

COHEN, M. (1987). A historical overview of the phenomenologic movement. *Image, Journal of Nursing Scholarship*. **19** (1), 31-34.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd Edition). Hillsdale, NJ: Lawrence Earlbaum Associates.

COLEMAN, E., RENFRO, R., et al. (1996). Reliability of the mantial dynamic mode of the Baltimore Therapeutic Equipment work simulator. *Journal of Hand Therapy*, **9**, 213-237.

COLLEGE OF OCCUPATIONAL THERAPISTS. (2007). *Professional standards for occupational therapy practice.* London: College of Occupational Therapists.

COOK, C., HEATH, F., THOMPSON, R. L. (2000). A meta-analysis of response rates in Web- or Internet-based surveys. *Educational and Psychological Measurement*, **60**, 821–836.

CONSTAND, M., AND MACDERMID, J. (2013). Effect of neck pain on reaching overhead and reading: a case-control study of long and short neck flexion. *BMC Sports Science, Medicine, and Rehabilitation*, **5**, 21.

CONNELL, L. A. and TYSON, S. F. (2012). Clinical reality of measuring upper-limb ability in neurologic conditions: A systematic review. *Arch Phys Med Rehabil*, **93**, 221-228.

COTE, P., CASSIDY, J., CARROLL, L., KRISTMAN, V. (2004). The annual incidence and course of neck pain in the general population: a population-based cohort study. *Pain*, **3**, 267-273.

COTE, P., van der VELDE, G., CASSIDY, J. D., et al. (2008). The burden and determinants of neck pain in workers: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and its associated disorders. *Spine*, **33**, S60–74.

COUPER, M. P. (2000). Web surveys – A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464–494.

COUPER, M. P., TRAUGOTT, M. W., LAMIAS, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, **65**, 230–253.

COUPER, M. P., TOURANGEAU, R., CONRAD, F. G., CRAWFORD, S. D. (2004). What they see is what we get – Response options for web surveys. *Social Science Computer Review*, **22**, 111–127.

COUPER, M. P., CONRAD, F. G., TOURANGEAU, R. (2007). Visual context effects in web surveys. *Public Opinion Quarterly*, **71**, 623–634.

CRAWFORD, S. D., MCCABE, S. E., POPE, D. (2005). Applying Web-based survey design standards. *Journal of Prevention and Intervention in the Community*, **29**, 43–66.

CROFT, P., LEWIS, M., et al. (2001) Risk factors for neck pain: a longitudinal study in the general population. *Pain,* **93**, 317-325.

CROTTY, M. (1998). *The Foundations of Social Research*. London: Sage.

CSP. (2005). *Core standards of physiotherapy practice*. London: Chartered Society of Physiotherapy.

CSP. (2012). *Quality Assurance Standards*: *Audit tool*. London: The Chartered Society of Physiotherapy. Available from: www.csp.org.uk/publications/quality-assurance-standards-audit-tool http://www.csp.org.uk/publications/quality-assurance-standards

CURB, J. D., CERIA-ULEP, C. D., et al. (2006). Performance-based measures of physical function for high-function populations. *Journal of the American Geriatrics Society*, **54** (5), 737-42.

DARNE, R. (2010). *Construct validity of single arm military press (SAMP) test in non-patient female with neck pain*. Sheffield: Sheffield Hallam University.

DAFFNER, S., HILIBRAND, A., HANSCOM, B., BRISLIN, B., VACCARO, A., ALBERT T. (2003). Impact of neck and arm pain on overall health status**.** *Spine*, **28,** 2030-2035.

de KONING, C. H., van den HEUVEL, S. P., STAAL, J. B., SMITS-ENGELSMAN, B. C., HENDRIKS, E. J. (2008). Clinimetric evaluation of active range of motion measures in patients with non-specific neck pain: a systematic review. *Eur Spine*, **17** (7), 905-921.

de VET, H. C. W., TERWEE, C. B., MOKKINK, L., KNOL, D. (2011). *Measurement in medicine: a practical guide*. Cambridge: CAMBRIDGE University Press.

de VET, H. C. W., TERWEE, B., KNOL, D. L., BOUTER, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, **59**, 1033-39.

de VET, H. C. W., ADER, H. J., TERWEE, C. B., et al. (2005). Are factor analytical techniques appropriately used in the validation of health status questionnaires? A systematic review on the quality of factor analyses of the SF-36. *Qual Life Res*,**14**, 1203-18.

DILLMAN, D. A., TORTORA, R. D., BOWKER, D. (1998). Principles for constructing web surveys. *Paper presented at the SESRC Technical Report* 90-50, Pullman, Washington.

DILLMAN, D. A. (2000). Mail and Internet surveys: The tailored design method. (2nd Edition) New York: John Wiley.

DILLMAN, D. A. (2007). *Mail and Internet Surveys: The tailored design method. update with new Internet, visual, and mixed-mode guide.* (2nd Edition) New York, NY: John Wiley & Sons.

DILLMAN, D. A. and SMYTH, J. D. (2007). Design effects in the transition to Web-based surveys. *American Journal of Preventive Medicine*, **32** (5S), 90–96.

DONOVAN, J. L., FRANKEL, S. J., EYLES, J. D. (1993). Assessing the need for health status measures. *J Epidemiol Community Health*, **47** (2),158-62.

DONOVAN, P. J. and PAULOS, L. E. (1995). Common injuries of the shoulder. Diagnosis and treatment. *Western Journal of Medicine*, **163**, 351-359.

DOMHOLDT, E. (2000). *Physical Therapy Research: Principles and Applications*. (2nd Edition) Philadelphia: PA Saunders.

DUCKWORTH, M. (1999). Outcome measurements, types, selecting instruments, difficulties, typology. *Physiotherapy*, **85**, 21-27.

DUNCKLEY, M., ASPINAL, F., ADDINGTON-HALL, J. M., HUGHES, R., HIGGINSON, I. J. (2005). A research study to identify facilitators and barriers to outcome measure implementation. *Int J Palliat Nurs*, **11** (5), 218-5.

DVIR, Z. and PRUSHANSKY, T. (2008). Cervical muscles strength testing: methods and clinical implications. *J Manipulative Physiol Ther*, **31** (7), 518-24.

ERICKSON, P., TAEUBER, R. C., SCOTT, J. (1995). Operational aspects of quality-of-life assessment: choosing the right instrument: review article. *PharmacoEconomics*, **7**, 39–48.

ERIKSEN, W., NATVIG, B., KNARDAHL, S., BRUUSGAARD, D. (1999). Job characteristics as predictors of neck pain. A 4-year prospective study. *Journal of Occupational and Environmental Medicine/American College of Occupational and Environmental Medicine*, **41** (10), 893-902.

EZZO, J. B., HARALDSSON, G., et al. (2007). Massage for mechanical neck disorders: a systematic review [with consumer summary]. *Spine,* **32** (3), 353-362.

FALLA, D., BILENKIJ, G., JULL, G. (2004). Patients with chronic neck pain demonstrate altered patterns of muscle activation during performance of a functional upper limb task. *International Journal of Clinical Practice*, **29**, 1436-1440.

FAN, Z., SMITH, C., SILVERSTEIN, B. (2008). Assessing validity of the QuickDASH and SF-12 as surveillance tools among workers with neck or upper extremity musculoskeletal disorders. *J HAND THER*, **21**, 354-365.

FAN, Z., SMITH, C., SILVERSTEIN, B. (2011). Responsiveness of the QuickDASH and SF-12 in workers with neck or upper extremity musculoskeletal disorders: one-year follow-up. *J Occup Rehabil*, **21**, 234-243.

FAN, W. and YAN, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computer in Human Behavior*, **26**, 132-139.

FAYAD, F., LEFEVRE-COLAU, M. M., et al. (2009). Reliability, validity and responsiveness of the French version of the questionnaire Quick Disability of the Arm, Shoulder and Hand in shoulder disorders. *Manual Therapy*, **14** (2), 206-212.

FEJER, R., KYVIK, K., HARTVIGSEN, J. (2006). The prevalence of neck pain in the world population: a systematic review of the literature. *Eur Spine J*, **15**, 834-848.

FELEUS, A., BIERMA-ZEINSTRA, S. M., MIEDEMA, H. S., et al. (2008). Incidence of nontraumatic complaints of arm, neck and shoulder in general practie. *Man Ther*, **13**, 426-33.

FEENY, D. H. and TORRANCE, G. W. (1989). Incorporating utility-based quality-of-life assessment measures in clinical trials. Two examples. *Med Care*, 27, S190-S204.

FITZPATRICK, R., DAVEY, C., BUXTON, M. J., JONES, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*, **2** (14), 1-74.

FRANK, A., SOUZA, L., FRANK, C. (2005). Neck pain and disability: A cross-sectional survey of the demographic and clinical characteristics of neck pain seen in a rheumatology clinic. *International Journal of Clinical Practice,* **59**, 173-182

FREDIN, K. and LORAS, H. (2017). Manual therapy, exercise therapy or combined treatment in the management of adult neck pain - A systematic review and meta-analysis. *Musculoskeletal Science Practice*, **31**, 62-71.

FRICKER, R. D. and SCHONLAU, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods*, **14**, 347–367.

FOSTER, N., THOMPSON, K., BAXTER, G., ALLEN, J. (1999). Management of nonspecific low back pain by physiotherapists in Britain and Ireland. A descriptive questionnaire of current clinical practice. *Spine*, **24**, 1332-1342.

FRANCHIGNONI, F., FERRIERO, G., et al. (2011). Psychometric properties of QuickDASH-A classical test theory and Rasch analysis study. *Manual Therapy*, **16** (2), 177-182.

FREBURGER, J. K., CAREY, T. S., HOLMES, G. M. (2005). Management of back and neck pain: who seeks care from physical therapists? *Phys Ther*, **85**, 872–86.

FU, L., LI, J., WU, W. (2009). Randomised controlled trials of acupuncture for neck pain: Systematic review and meta-analysis. *THE JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE*, **15**, 133-145.

FURLAN, D., PENNICK, V., BOMBARDIER, C., van TULDER, M. (2009). 2009 Update methods guidelines for systematic review in the Cochrane Back Review Group. *Spine*, **34** (18), 1929-1941.

GABEL, C. P., YELLAND, M., et al. (2009). A modified QuickDASH-9 provides a valid outcome instrument for upper limb function. *BMC musculoskeletal disorders*, **10** (1), 161.

GARLAND, A., MARC KRUSE, B. A., AARONS, G. (2003). Clinicians and outcome measurement: what's the use? *Journal of Behavioral Health Services & Research*, **30** (4) 393-405.

GEMMELL, H. and MILLER, P. (2006). Comparative effectiveness of manipulation, mobilisation and the activator instrument in treatment of non-specific neck pain: a systematic review. *Chiropr Osteopat*, **14**, 7.

GIRAUDEAU, B., MARY, J. Y. (2001). Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in Medicine*, **20**, 3205-14.

GORSKI, J. M. and SCHWARTZ, L. H. (2003). Shoulder impingement presenting as neck pain. *The journal of bone and joint surgery*, **85** (4), 635-638.

GRAHAM, N., GROSS, A., GOLDSMITH, C. (2006). Mechanical traction for mechanical neck disorders: A systematic review. *J Rehabil Med*, **38**, 145-152.

GRAHAM N., GROSS, A., GOLDSMITH, C., KLABER MOFFETT, J., HAINES, T., BURNIE, S., et al. (2008). Mechanical traction for neck pain with or without radiculopathy. *The Cochrane Library*, **3**.

GRAY, J. (1997). *Evidence-based Healthcare. How to Make Health Policy and Management Decisions*. New York: Churchill Livingstone.

GREENHALGH, T. (2010). *How to read a paper.* (4th edition) UK, Chichester: Wiley-Black well.

GREENSLADE, J. R., MEHTA, R. L., BELWARD, P., et al. (2004). Dash and Boston questionnaire assessment of carpal tunnel syndrome outcome: what is the responsiveness of an outcome questionnaire? *J Hand Surg* [Br], **29**, 159–64.

GROSS, A., FORGET, M. S. T., GEORGE K, FRASER, M., GRAHAM, N., PERRY, L., et al. (2012). Patient education for neck pain (Review). *The Cochrane Library*; **3.**

GROSS, A., MILLER, J., D'SYLVA, J., Burnie, S., Goldsmith, C., Graham, N., et al. (2010). Manipulation or mobilisation for neck pain: A Cochrane review. *Manual Therapy*, **15**, 315-333.

GROSS, A. R., GOLDSMITH, C., et al. (2007). Conservative management of mechanical neck disorders: a systematic review. *The Journal of Rheumatology*, **34** (5), 1083-1102.

GROSS, A., HOVING, J., HAINES, T. (2004). A Cochrane review of manipulation and mobilization for mechanical neck disorders. *Spine*, **29**, 1541-1548.

GROSS, A., KAY, T., HONDRAS, M., GOLDSMITH, C., HAINES, T., PELOSO, P., et al. (2002). Manual therapy for mechanical neck disorders: A systematic review. *Man Ther*, **7**, 131-149.

GUMMESSON, C., WARD, M. M., et al. (2006). The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC Musculoskeletal Disorders*, **7** (1), 44.

GUYATT, G., WALTER, S., NORMAN, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*, **40**, 171–8.

HAAS, F., HUBMER, M., et al. (2011). Long-term subjective and functional evaluation after thumb replantation with special attention to the QuickDASH questionnaire and a specially designed trauma score called modified mayo score. *Journal of Trauma-Injury, Infection, and Critical Care*, **71** (2), 460-466.

HAIGH, R., TENNANT, A., BIERING-SORENSEN, F., et al (2001). The use of outcome measures in physical medicine and rehabilitation within Europe. *J Rehabil Med*, **33**, 273–278.

HAINES, T., FROSS, A., BURNIE, S., GOLDSMITH, C., PERRY, L., GRAHAM, N., et al. (2009). A Cochrane review of patient education for neck pain. *Spine*, **9**, 859-871.

HALDEMAN, S., CARROLL, L., et al. (2008). The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders: executive summary. *Spine*, (Phila Pa 1976) **33** (4 Suppl), S5-7.

HALDEMAN, S., CARROLL, L., CASSIDY, J. D. (2010). Findings from the bone and joint decade 2000 to 2010 task force on neck pain and its associated disorders. *J Occup Environ Med*, **52**, 424-427.

HAKKINEN, A., KAUTIAINEN, H., et al. (2008). Strength training and stretching versus stretching only in the treatment of patients with chronic neck pain: a randomized one-year follow-up study. *Clin Rehabil*, **22** (7), 592-600.

HAKKINEN, A., SALO, P., et al. (2007). Effect of manual therapy and stretching on neck muscle strength and mobility in chronic neck pain. *J Rehabil Med* **39** (7), 575-579.

HAMMOND, R. (2000). Evaluation of physiotherapy by measuring the outcome. *Physiotherapy*, **86** (4), 170-172.

HARALDSSON, B., GROSS, A., et al. (2006). Massage for mechanical neck disorders. *Cochrane Database of Systematic Reviews* DOI: 10.1002/14651858.CD004871.pub3.

HASSON, D. and ARNETZ, B. (2005). Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. *International Electronic Journal of Health Education*, **8**, 178-192.

HAYWOOD, K. L., TUTTON, E., STANISZEWSKA, S. (2013). Patient-reported outcome measures (PROMs) and fragility hip fracture: A systematic review of quality and acceptability. *Quality of Life Research*, **22**, 2077.

HAYWOOD, K. L., COLLIN, S. M., CRAWLEY, E. (2014) Assessing severity of illness and outcomes of treatment in children with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review of patient-reported outcome measures (PROMs). *Child Care Health Dev*, **40**, 806–24.

HCPC. (2016). *Annual report and account 215-16 and gender breakdown*. http://www.hpc-uk.org/assets/documents/100050E6Annualreportandaccounts2015-16.pdf
http://www.hcpc-uk.co.uk/assets/documents/10005135Registrants-by-gender-England-July-2016.pdf
http://www.hcpc-uk.co.uk/assets/documents/10005139Registrants-by-gender-Scotland-July-2016.pdf
http://www.hcpc-uk.co.uk/assets/documents/10005138Registrants-by-gender-NIreland-July-2016.pdf
http://www.hcpc-uk.co.uk/assets/documents/1000513BRegistrants-by-gender-Wales-July-2016.pdf

HCPC. (2013). *Standard of proficiency*. http://www.hpc-uk.org/assets/documents/10000DBCStandards_of_Proficiency_Physiotherapists.pdf

HEATON, J. and BAMFORD, C. (2001). Assessing the outcomes of equipment and adaptations: Issues and approaches. *British Journal of Occupational Therapy*, **64** (7), 346-356.

HERZOG, W. (2010). The Biomechanics of Spinal Manipulation. *Journal of Bodyworks and Movement Therapies*. **14**, 280-286.

HENLEY, L. and FRANK, D. (2006). Reporting ethical protection in physical therapy research. *PHYS THER*, **86**, 499-509.

HOGG-JOHNSON, S., van der VELDE, G., et al. (2008). The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. *Spine,* **33**(4), S39–51.

HORN, K. K., JENNINGS, S., RICHARDSON, G., VLIET, D. V., HEFFORD, C., ABBOTT, J. H. (2012). The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. *J Orthop Sports Phys Ther*, **42** (1), 30-42.

HOOZEMANS, M., van der BEEK, A., FRINGS-DRESEN, van der WOULDE, L., van DIJK, F. (2002). Pushing and pulling in association with low back and shoulder complaints. *Occupational & Environmental Medicine*, **59** (10), 696-702.

HOVING, J., KOES, B., de VET, H. C. W., van der WINDT, D., ASSENDELFT, W., MAMEREN, H., et al. (2002) Manual therapy, physical therapy, or continued care by a general practitioner for patients with neck pain. *Ann Intern Med*, **136**, 713-722.

HOVING, J., de VET, H., TWISK, J., DEVILLE, W., van der WINDT, D., KOES, B.,

BOUTER, L. (2004). Prognostic factors for neck pain in general practice. Pain, **110** (3), 639-645.

HOY, D., PROTANI, M., De, R., BUCHBINDER, R. (2010). The epidemiology of neck pain. *Best Practice & Research Clinical Rheumatology*, **24**, 783-792.

HOY, D., MARCH, L., WOOLF, A., et al. (2014). The global burden of neck pain: estimates from the global burden of disease 2010 study. *Ann Rheum Dis,* **73**, 1309-1315.

HUDAK, P., AMADIO, P., BOMBARDIER, C. (1996). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The upper extremity collaorative group (UECG). *Am J Ind Med*, **29**, 602-608.

HUGHES, D. J., ROWE, M., BATEY, M., LEE, A. (2012). A tale of two sites: Twitter vs. Facebook and personality predictors of social media usage. *Computers in Human Behavior*, **28**, 561-569.

HUISSTEDE, B., MIEDEMA, H., VERHAGEN, A., KOES, B., VERHAAR, J. (2007). Multidisciplinary consensus on the terminology and classification of complaints of the arm, neck/or shoulder. *Occup Environ Med*, **64**, 313-319.

HUISSTEDE, B., BIERMA-ZEINSTRA, S., KOES, B., VERHAAR, J. (2006). Incidence and prevalence of upper extremity musculoskeletal disorders: A systematic appraisal of the literature. *BMC Musculoskeletal Disord*, **7**, 7.

HUISSTEDE, B., FELEUS, A., BIERMA-ZEINSTRA, S., VERHAAR, J., KOES, B. (2009). Is the disability of arm, shoulder, and hand questionnaire (DASH) also valid and responsive in patients with neck complaints. *Spine*, **34**, 130-138.

HURWITZ, E. L., CARRAGEE, E. J., van der VELDE, G., CARROLL, L. J., NORDIN, M., GUZMAN, J., et al. (2008). Treatment of neck pain: noninvasive interventions. Results of the Bone and Joint Decade 2000 to 2010 Task Force on Neck Pain and its Associated Disorders [with consumer summary]. *Spine*. **33** (4 Suppl), S123-S52.

ITOH, K., KATSUMI, Y., et al. (2007). Randomised trial of trigger point acupuncture compared with other acupuncture for treatment of chronic neck pain. *Complement Ther Med***, 15** (3), 172-179.

JAIN, V. (2010). *Inter- and intra-rater reliability of single arm military press (SAMP) test in a female non-patient with neck pain*. Sheffield: Sheffield Hallam University.

JAESCHKE, R., SINGER, J., GUYATT, G. H. (1990). A comparison of seven-point and visual analogue scales. Data from a randomized trial. *Control Clin Trials*, **11**, 43–51.

JAESCHKE, R., SINGER, J., GUYATT, G. H. (1990). A comparison of seven-point and visual analogue scales. Data from a randomized trial. *Control Clin Trials*, **11** (1), 43-51.

JEAVONS, A. (1998). Ethology and the Web: Observing Respondent Behaviour in Web Survey. *In proceedings of the Worldwide Internet Conference (CD)*. London: ESOMAR.

JETTE, D., HALBERT, J., IVERSON, C., MICELI, E., SHAH, P. (2009). Use of standardised outcome measures in physical therapist practice: perception and applications. *Phys Ther*, **89**, 125-135.

JULL, G., STERLING, M., FALLA, D., TRELEAVEN, J., O'LEARY, S. (2008). *Whiplash, head and neck pain: Research-based directions for physical therapies*. London: Churchill Livingstone.

KAY, T., GROSS, A., GOLDSMITH, C., RUTHERFORD, S., VOTH, S., HOVING, J., et al. (2012). Exercise for mechanical neck disorders (review). *The Cochrane Library*, **8**.

KAY, T., GROSS, A., GOLDSMITH, C. (2005). Exercises for mechanical neck disorders. *Cochrane database Syst Rev*, **3**, CD004250.

KESSLER, R. C. and MROCZEK, D. K. (1995). Measuring the effects of medical interventions. *Med Care*, **33**, AS109–19.

KLEE, R. (1997). *Introduction to the philosophy of science—cutting nature at its seams*. New York, Oxford University Press.

KRAMER, A. M. and HOLTHAUS, D. (2006). Uniform patient assessment for post-acute care. Available at: https://aspe.hhs.gov/system/files/pdf/75046/strokePAC.pdf

KROELING, P., GROSS, A., GOLDSMITH, C. (2005). A Cochrane review of electrotherapy for mechanical neck disorders. *Spine*, **30**, 641-648.

KROELING, P., GROSS, A., GRAHAM, N., BURNIE, S., SZETO, G., GOLDSMITH, C., et al. (2013). Electrotherapy for neck pain (review). *The Cochrane Library*, **8**.

KULKARNI, M. (2010). *Inter- and intra-rater reliability of the single arm military press (SAMP) test in females with neck pain*. Sheffield: Sheffield Hallam University.

KUMTA, P., MACDERMID, J., MEHTA, S., STRATFORD, P. (2012). The FIT-HaNSA Demonstrates Reliability and Convergent Validity of Functional Performance in Patients with Shoulder Disorders. *journal of orthopaedic & sports physical therapy*, **5,** 42.

KWAK, S., NIEDERKLEIN, R., TARCHA, R., HUGHES, C. (2005). Relationship between active cervical range of motion and perceived neck disability in community dwelling elderly individuals. *Journal of Geriatric Physical Therapy,* **28**, 54-56.

KYTE, D., CALVERT, M., van der WEES, P., HOVE, R., TOLAN, S., HILL, J. (2015). An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy,* **101**, 119-125.

LANSKY, D., BUTLER, J. B., WALLER, F. T. (1992). Using health status measures in the hospital setting: from acute care to 'outcomes management'. *Med Care*, **30**, MS57–73.

LANDIS, J. R. and KOCH, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33** (1), 159-74.

IANDSHEERE, G. (1988). *Research perspectives. In: Keeves J, ed. Educational Research Methodology and Measurement.* New York, NY: Pergamon Press Inc.

LATHAM, N. K., MEHTA, V., NGUYEN, A. M., JETTE, A. M., et al. (2008). Performance-based or self-reported measures of physical function: Which should be used in clinical trials of hip fracture patients? *Arch Physical Medicine and Rehabilitation*, **89** (11), 2146-2155.

LEE, C. E., SIMMONDS, M. J., NOVY, D. M., JONES, S. C. (2000). A comparison of self-report and clinician measured physical function among patients with low back pain. *Arch Phys Med Rehabil,* **82**, 227–31.

LITTLEWOOD, C. and MAY, S. (2013). *Understanding physiotherapy research*. (1st Edition) Cambridge: Scholars Publishing.

LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GOTZSCHE, P. C., IOANNIDIS, J. P., CLARKE, M., DEVEREAUX, P. J., KLEIJNEN, J., MOHER, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, **339**, b2700.

LIANG, H. W., WANG, H. K., YAO, G., HORNG, Y. S., HOU, S. M. (2004). Psychometric evaluation of the Taiwan version of the Disability of the Arm, Shoulder, and Hand (DASH) questionnaire. *J Formos Med Assoc*, **103**, 773–9.

LIANG, Z., ZHU, X., YANG, X., FU, W., LU, A. (2011). Assessment of traditional acupuncture therapy for chronic neck pain: a pilot randomised controlled study. *Complementary Therapies in Medicine*, **195**, 26-32.

LINTON, S. (2000). A review of psychological risk factors in back and neck pain. *Spine*, **25**, 1148–56.

LOMOND, K. and COTE, J. (2009). Shoulder functional assessments in persons with chronic neck/shoulder pain and healthy subjects: Reliability and effects of movement repetition. *Work*, **38**, 169-180.

LUIME, J., KOES, B., MIEDEM, H., VERHAAR, J., BURDORF, A. (2005). High incidence and recurrence of shoulder and neck pain in nursing home employees was demonstrated during a 2-year follow-up. *J Clin Epidemiol*, **58**, 407-413.

LUNDBLAD, I., ELERT, J., GERDLE, B. (1999). Randomised controlled trial of physiotherapy and Feldenkrais interventions in female workers with neck-shoulder complaints. *Journal of Occupational Rehabilitation*, **9**, 179-194.

MACAULAY, J., CAMERON, M., VAUGHAN, B. (2007). The effectiveness of manual therapy for neck pain: a systematic review of the literature. *Physical Therapy Reviews*, **12** (3), 261-7.

MACDERMID, J., WALTON, D., COTE, P., SANTAGUIDA, P., GROSS, A., CARLESSO, L. (ICON) (2013). Use of outcome measures in managing neck pain: an international multidisciplinary survey. *The Open Orthopaedics Journal*, **7**, 506-520.

MACDERMID, J. C., WALTON, D. M., AVERY, S., et al. (2009). Measurement properties of the neck disability index: a systematic review. *J Orthop Sports Phys Ther*, **39** (5), 400-17.

MACDERMID, J., GHOBRIAL, M., QUIRION, K., ST-AMOUR, M., TSUI, T., et al. (2007). Validation of a new test that assesses functional performance of the upper extremity and neck (FIT-HaNSA) in patients with shoulder pathology. *BMC Musculoskeletal Disorders*, **8**, 42.

MANFREDA, K. L., BOSNJAK, M., BERZELAK, J., HAAS, I., VEHOVAR, V. (2008). Web surveys versus other survey modes. *International Journal of Market Research*, **50**, 79–104.

MARDANI-KIVI, M., KARIMI-MOBARAKEH, M., et al. (2013). The effects of corticosteroid injection versus local anaesthetic injection in the treatment of lateral epicondylitis: a randomised single-blinded clinical trial. *Archives of Orthopaedic and Trauma Surgery*, **133** (6), 757-763.

MATHESON, L. N., MELHORN, J. M., et al. (2006). Reliability of a visual analog version of the QuickDASH. *The Journal of Bone & Joint Surgery*, **88** (8), 1782-1787.

MAYER, M. M. (2012). Seeking what matters. *The Patient-Patient-Centered Outcomes Research*, **5**, 71-74.

MCCORMACK, H., M., HORNE, D. J., SHEATHER, S. (1988). Clinical applications of visual analogue scales: a critical review. *Psychol Med*, **18**, 1007–19.

MCDOWELL, I. (2006). *The theoretical and technical foundation of health measurement. In: Measuring health: a guide to rating scales and questionnaires*. 3rd ed. New York: Oxford University Press.

MCGRAW, K. O. and WONG, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol Methods*, **1**, 30–46.

MCLEAN, S., MAY, S., KLABER-MOFFETT, J., SHARP, D., GARDINER, E. (2007). Prognostic factors for progressive non-specific neck pain. *Physical Therapy Reviews*, **12**, 207-220.

MCLEAN, S. (2007). *Conservative management of non-specific neck pain: Effectiveness of treatment, predictors of treatment outcome and upper limb disability*. PhD dissertation, University of Hull; Hull.

MCLEAN, S., TAYLOR, J., et al. (2010a). Measuring upper limb disability in non-specific neck pain: a clinical performance measure. *International Journal of Physical*

Therapy and Rehabilitation, **1**, 44-52

MCLEAN, S., MAY, S., KLABER-MOFFETT, J., SHARP, D., GARDINER, E. (2010b). Risk factors for the onset of non-specific neck pain: a systematic review. *J Epidemiol Community Health*, **64**, 565-572.

MCLEAN, S., MOFFETT, J., SHARP, D., GARDINER, E. (2011). An investigation to determine the association between neck pain and upper limb disability for patients with non-specific neck pain: A secondary analysis. *Manual Therapy*, **1**, 1-6.

MCLEAN, S. M., MOFFETT, J. A., SHARP, D. M., GARDINER, E. (2013). A randomised controlled trial comparing graded exercise treatment and usual physiotherapy for patients with non-specific neck pain. *Manual Therapy*, **18** (3), 199-205.

MEHTA, S., MACDERMID, J., CARLESSO, L., MCPHEE, C. (2010). Concurrent validation of the DASH and the QuickDASH in comparison to neck-specific scales in patients with neck pain. *Spine*, **35**, 2150-2156.

MERSKEY, H. and BOGDUK, N. (1994). *Classification of Chronic Pain*. (2nd Edition) IASP Seattle: Task Force on Taxonomy IASP Press.

MINTKEN, P. E., GLYNN, P., et al. (2009). Psychometric properties of the shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. *Journal of Shoulder and Elbow Surgery*, **18** (6), 920-926.

MILLER, J., GROSS, A., D'SYLVA, J., BURNIE, S., GOLDSMITH, C., GRAHAM, N. (2010). Manual therapy and exercise for neck pain: A systematic review. *Manual Therapy*, **15**, 334-354.

MOFFETT, J. and MCLEAN, S. (2006). The role of physiotherapy in the management of non-specific back pain and neck pain. *Rheumatology*, **45**, 371-378.

MOKKINK, L., TERWEE, C., et al. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, **19**, 539-549.

MOKKINK, L. B., TERWEE, C. B., et al. (2010b). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol, 63, 737-45.

MOKKINK, L. B., TERWEE, C. B., et al. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res*, **18**, 313-33.

MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. and GROUP, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, **339**, b2535.

MORADI, A., MENENDEZ, M. E., KACHOOEI, A. R., ISAKOV, A., RING, D. (2016). Update of the Quick DASH questionnaire to account for modern technology. *Hand*, **11**(4) 403-409.

MORGAN, G. A. and HARMON, R. J. (1999). Clinicians' guide to research methods and statistics: Sampling and external validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, **38** (8), 1051-1053.

NAKAMOTO, H., OSHIMA, Y., et al. (2014). Usefulness of QuickDASH in patients with cervical laminoplasty. *J Orthop Sci*, **19** (2), 218-222

NELSON, E. C., LANDGRAF, J. M., HAYS, R. D., WASSON, J. H., KIRK, J. W. (1990). The functional status of patients. How can it be measured in physicians' offices? *Med Care*, **28**, 1111–26.

NELSON, E., BERWICK, D. (1989). The measurement of health status in clinical practice. *Med Care*, **27**, S77–S90.

NHS. (2010). *White paper Equity and Excellence: Liberation the NHS.* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf

NIEKEL, M. C., LINDENHOVIUS, A. L., et al. (2009). Correlation of DASH and QuickDASH with measures of psychological distress. *The Journal of Hand Surgery*, **34** (8), 1499-1505.

NORDENFELT, L. (1994). *Concepts and measurement of quality of life in health care*. Dordrecht: Kluwer Academic Publishers.

NORDIN, M., CARRAGEE, E. J., et al. (2008). Assessment of neck pain and its associated disorders: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine* (Phila Pa 1976), **33** (4 Suppl), S101-22.

OFFENBACHER, M., EWERT, T., SANGHA, O., STUCKI, G. (2003). Validation of a German version of the 'Disabilities of Arm, Shoulder and Hand' (DASH-G). *Z Rheumatol*, **62**, 168–77.

OFFICE FOR NATIONAL STATISTICS (2016). Internet users in the UK: Statistical bulletin.www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2016.

OLSON, S., O'CONNOR, D., BIRMINGHAM, G., BROMAN, P., HERRERA, L. (2000). Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. *The Journal of Orthopaedic and Sports Physical Therapy*, **30**, 13-20.

O'LEARY, S., JULL, G., et al. (2007). Specificity in retraining craniocervical flexor muscle performance. *Journal of Orthopaedic and Sports Physical Therapy*, **37** (1), 3-9.

O'RIORDAN, C., CLIFFORD, A., de VEN, P., NELSON, J. (2014). Chronic neck pain and exercise interventions: frequency, intensity, time, and type principle. *Archives of Physical Medicine and Rehabilitation*, **95**, 770-783.

OSBORN, W. and JULL, G. (2013). Patients with non-specific neck disorders commonly report upper limb disability**.** *Manual Therapy*, **18**, 492-497.

PATEKAR, P. (2010). *Clinical utility of single arm military press (SAMP) test in non-patient females with neck pain*. Sheffield: Sheffield Hallam University.

PATEL, K., GROSS, A., GRAHAM, N., GOLDSMITH, C., EZZO, J., MORIEN, A., et al. (2012). Massage for mechanical neck disorders (Review). *The Cochrane Library*, **9.**

PEREZ, H. I., PEREZ, J. L., MARTINZ, A. G., et al. (2014) Is one better than another? A randomized clinical trial of manual therapy for patients with chronic neck pain. Manual Therapy, **19**, 215-221

PETTY, N. (2011). *Neuromusculoskeletal examination and assessment: a handbook for therapists*. (4th edition) Edinburgh: ELSEVIER.

PIETROBON, R., COEYTAUX, R. R., CAREY, T. S., RICHARDSON, W. J., DEVELLIS, R. F. (2002). Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine*, **27** (5), 515-22.

PHILLIPS D. (1987) *Philosophy, Science, and Social Inquiry*. New York, NY: Pergamon Press Inc.

PINHEIRO, L. C., CALLAHAN, L. F., CLEVELAND, R. J., EDWARDS, L. J., REEVE, B. B. (2016). The Performance and Association Between Patient-reported and Performance-based Measures of Physical Functioning in Research on Individuals with Arthritis. *J Rheumatology,* **43**, 131-137.

POLSTON, D. (2007). Cervical radiculopathy**.** *Neurol. Clin*, **25**, 373-385.

POITRAS, S., LOISEL, P., PRINCE, F., LEMAIRE, J. (2000). Disability measurement in persons with back pain: a validity study of spinal range of motion and velocity. *Archives of Physical Medicine and Rehabilitation*, **81**, 1394-1400.

PORTNEY, L. G. and WATKINS, M. P. (2009). *Foundations of clinical research: Applications to practice, Upper Saddle River*, NJ, Pearson Prentice Hall.

POLSON, K., REID, D., et al. (2010). Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Manual Therapy*, **15** (4), 404-407.

QUATMAN-YATES, C., GUPTA R., et al. (2013). Internal consistency and validity of the QuickDASH instrument for upper extremity injuries in older children. *J Pediatr Orthop*, **33** (8), 838-842.

RASMUSSEN, C., LEBOEUF-YDE, C., HESTBAEK, L., et al. (2008). Poor outcome in patients with spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in the secondary care sector. *Scand J Rheumatol*, **37**, 462-8

RAVEN, E. E., HAVERKAMP, D., SIEREVELT, I. N., et al. (2008). Construct validity and reliability of the disability of arm, shoulder and hand questionnaire for upper extremity complaints in rheumatoid arthritis. *J Rheumatol*, **35**, 2334–8.

READ, J. L., QUINN, R. J., HOEFER, M. A. (1987). Measuring overall health: an evaluation of three important approaches. *J Chronic Dis*, **40** (1), 7S–26S.

REEVE, B. B., WYRWICH, K., et al. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centred outcomes and comparative effectiveness research. *Qual Life Res*, **22** (8),1889-905

REMINGTON, M., TYRER, P., NEWSON, J., SMITH, J., CICCHETTI, D. V. (1979). Comparative reliability of categorical and analogue rating scales in the assessment of psychiatric symptomatology. *Psychol Med*, **9**, 765–70.

RENEMAN, M. F., JORRITSMA, W., SCHELLEKENS, J. M. H., et al. (2002). Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *J Occup Rehabil*, **12**, 119–29.

RHEE, J., YOON, T., RIEW, K. (2007). Cervical Radiculopathy. *The Journal of the American Academy of Orthopaedic Surgeons,* **15**, 486-494.

ROSE, M., BJORNER, J. B., BECKER, J., FRIES, J. F., WARE, J. E. (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*, **61**, 17-33.

RUDD, A., PEARSON, M., GEORGIOU, A. (2000). *Measuring Clinical Outcome in stroke (acute care): Clinical effectiveness and evaluation unit*. London: Royal College of Physicians.

SACKETT, D. L., ROSENBERGWMC, GRAYJAM, RICHARDSON, W. S. (1996). Evidence based medicine: What it is and what it isn't. *Br MedJ*, **312**, 71-72.

SACKETT, D., RICHARDSON, W. S., ROSENBERG, W., HAYNES, R. B. (1997). *How to practice and teach evidence-based medicine*. Churchill Livingstone.

SAHIN, N., KARATAS, O., OZKAYA, M., et al. (2008). Demographics features, clinical findings and functional status in a group of subjects with cervical myofascial pain syndrome. *Agri,* **20**, 14–9.

SARIGIOVANNIS, P. and HOLLINS, B. (2005). Effectiveness of manual therapy in the treatment of non-specific neck pain: a review. *Physical Therapy Reviews*, **10** (1), 35-50.

SARIG-BAHAT, H. (2003). Evidence for exercise therapy in mechanical neck disorders. *Man Ther*, **8** (1), 10-20.

SCARABOTTOLO, C., PINTO, R., OLIVEIRA, C., ZANUTO, E., CARDOSO, J., CHRISTOFARO, D. (2017). Back and neck pain prevalence and their association with physical inactivity domains in adolescents. *Eur Spine J,* DOI 10.1007/s00586-017-5144-1.

SCOTT, P. J., HUSKISSON, E. C. (1977). Measurement of functional capacity with visual analogue scales. *Rheumatol Rehabil*, **16** (4), 257-259.

SCHELLINGERHOUT, J. M., HEYMANS, M. W., VERHAGEN, A. P., de VAT, H. C. W., KOES, B. W., TERWEE, C. B. (2011). Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Medical Research Methodology*, **11**, 87.

SCHELLINGERHOUT, J. M., VERHAGEN, A. P., HEYMANS, M. W., KOES, B. W., de VAT, H. C. W., TERWEE, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res*, **21**, 659-670.

SELBY, P. and ROBERTSON, B. (1987). Measurement of quality of life in patients with cancer. *Cancer Surv*, **6**, 521–43.

SHAVELSON, R. J. and WEBB, N. M. (1991). *Generalizability theory*. A primer. London: Sage Publications.

SHERMAN, K., COOK, A., WELLMAN, R., HAWKES, R., KAHN, J., DEYO, R., CHERKIN, D. (2014). Five-Week outcomes from a dosing trial of therapeutic massage for chronic neck pain. *Ann Fam Med*, **12**, 112-120.

SHEPARD, K., JENSEN, G., HACK, L., GWYER, J. (1993). Alternative approaches to research in physical therapy: positivism and phenomenology. *Physical Therapy,* **73**, 34/88-43/97.

SKEAT, J. and PERRY, A. (2008). Exploring the implementation and use of outcome measurement in practice: a qualitative study. *Int J Lang Commun Disord*, **43** (2), 110-25.

SILVA, A. G., PUNT, T. D., JOHNSON, M. I. (2010). Reliability and validity of head posture assessment by observation and a four-category scale. *Man Ther*, **15** (5), 490-5.

SILVERSTEIN, B., WELP, E., NELSON, N., KALAT, J. (1998). Claims incidence of work-related disorders of the upper extremities**.** Washington State, 1987 through 1995. *American Journal of Public Health*, **88**, 1827-1833.

SHAHEEN, A. A., OMAR, M. T., VERNON, H. (2013). Cross-cultural adaptation, reliability, and validity of the Arabic version of neck disability index in patients with neck pain. *Spin*, **38**, (10), E609-15.

SMEETS, R. J., et al. (2006). The association of physical deconditioning and chronic low back pain: A hypothesis-oriented systematic review. *Disability and rehabilitation*, **28** (11), 673-693.

SOOHOO, N. F., MCDONALD, A. P., SEILER, J. G., MCGILLIVARY, G. R. (2002). Evaluation of the construct validity of the DASH questionnaire by correlation to the SF-36. *J Hand Surg A*, **27**, 537—41.

SPRANGERS, M. A., CULL, A., BJORDAL, K., GROENVOLD, M., AARONSON, N. K. (1993). The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines for developing questionnaire modules. EORTC Study Group on Quality of Life. *Qual Life Res*, **2**, 287–95.

STERLING, M., KENARDY, J., JULL, G., VICENZINO, B. (2003). The development of psychological changes following whiplash injury. *Pain*, **106** (3), 481-9.

STRATFORD, P. W., KENNEDY, D. M., WOODHOUSE, L. J. (2006). Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. *Phys Ther*, 86, 1489–96.

STRATFORD, P. W., KENNEDY, D. M., MALY, M. R., MACINTYRE, N. J. (2010). Quantifying self-report measures' overestimation of mobility scores postarthroplasty. *Phys Ther*, **90**, 1288–96.

STREINER, D. L. and NORMAN, G. R. (2003). *Health measurement scales*. Oxford: Oxford University Press.

STOCK, S., LOISEL, P., DURAND, M., et al. (2003). L'indice d'impact de la douleur au cou et aux membres supérieurs sur la vie quotidienne (NULI: neck and upper limb index). *Études et Recherches*. R-355. Available at: http://www.irsst.qc.ca/media/documents/pubirsst/R-355.pdf

STANISZEWSKA, S., HAYWOOD, K. L., BRETT, J., TUTTON, L. (2012). Patient and public involvement in patient-reported outcome measures. *The Patient-Patient-Centred Outcomes Research*, **5**, 79-87.

SWENSON, R. (2003). Therapeutic modalities in the management of nonspecific neck pain. *Phys Med Rehabil Clin N Am*, **14**, 605-627.

TASHAKKORI, A. and TEDDLIE, C. (2010). *Sage Handbook of Mixed Methods in Social & Behavioral Research.* SAGE Publications Inc.

TERWEE, C. B., van der SLIKKE, R. M., van LUMMEL, R. C., BENINK, R. J., MEIJERS, W. G., de VET, H. C. W. (2006). Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. *J Clin Epidemiol*, **59**, 724–31.

TERWEE, C., BOT, S., de BOER, M., van der WINDT, D., KNOL, D., DEKKER, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, **60**, 34-42.

TERWEE, C., JANSMA, E., RIPHAGEN, I., de VET, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, **18**, 1115-1123.

TERWEE, C., MOKKINK, L., KNOL, D., OSTELO, R., BOUTER, L., de VET, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, **21**, 651-657.

TERWEE, C. B., PRINSEN, C. A. C., CHIAEOTTI, A., et al. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, (https://doi.org/10.1007/s11136-018-1829-0).

THOMAS, E., PEAT, G., HARRIS, L., WILKIE, R., CROFT, P. (2004). The prevalence of pain and pain interference in a general population of older adults: cross-sectional findings from the North Staffordshire Osteoarthritis Project (NorStOP). *Pain*, **110**, (1–2), 361–368.

TRINH, K., GROSS, A., GOLDSMITH, C. (2006). Acupuncture for neck disorders. *Cochrane database system review*, **3**.

TROSSMAN, P. B., SULESKI, K. B., LI, P. W. (1990). Test-retest reliability and day to day variability of an isometric grip strength test using the work simulator. *Occupational Therapy foumal of Research*, **10**, 266-279.

TROCHIM, W. (2002). The research methods knowledge base. Available online from https://www.researchgate.net/publication/243783609

TSAKITZIDIS, G., REMMEN, R., DANKAERTS, W., van ROYEN, P. (2013). Non-specific neck pain and evidence-based practice. *European Scientific Journal*, **9** (3) 1857-7431.

TOULASSIDHARANE, B. (2010). *Construct validity of the single arm military press (SAMP) test in a female non-patient population*. Sheffield: Sheffield Hallam University.

TURCHIN, D. C., BEATON, D. E., RICHARDS, R. R. (1998). Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg A*, **80**, 154—62.

van TULDER, M., FURLAN, A., BOMBARDIER, C., BOUTER, L. (Editorial Board of the Cochrane Collaboration Back Review Group) (2003). Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine*, **28** (12):1290e9.

VAS, J., PEREA-MILLA, E., et al. (2006). Efficacy and safety of acupuncture for chronic uncomplicated neck pain: a randomised controlled study. *Pain*, **126** (1-3), 245-255.

van DIJK, M., KOOT, H. M., SAAD, H. H., TIBBOEL, D., PASSCHIER, J. (2002). Observational visual analog scale in paediatric pain assessment: useful tool or good riddance? *Clin J Pain*, **18** (5), 310-316.

van DULMEN, S., van der WEES, P., BART STAAL, J., BRASPENNING, J., NIJHUIS-van der SANDEN, M. (2017). Patient reported outcome measures (PROMs) for goalsetting and outcome measurement in primary care physiotherapy, an explorative field study. *Physiotherapy,* **103**, 66-72.

VEEHOF, M. M., SLEEGERS, E. J., van VELDHOVEN, N. H., SCHUURMAN, A. H., van MEETEREN, N. L. (2002). Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther*, **15**, 347–354

VERNON, H. T., HUMPHREYS, B. K., HAGINO, C. A. (2005). A systematic review of conservative treatments for acute neck pain not due to whiplash. *Journal of Manipulative and Physiological Therapeutics*, **28** (6), 443-8

VERNON, H., HUMPHREYS, K., HAGINO, C. (2007). Chronic mechanical neck pain in adults treated by manual therapy: a systematic review of change scores in randomized clinical trials. *Journal of Manipulative and Physiological Therapeutics*, **30** (3), 215-27.

VERNON, H. and HUMPHREYS, B. K. (2007). Manual therapy for neck pain: an overview of randomized clinical trials and systematic reviews. *Europa MedicoPhysica* [Mediterraneal Journal of Physical and Rehabilitation Medicine], **43** (1), 91-118.

VERNON, H. and MIOR, S. (1991). The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther*, **14**, 409–15.

VICKERS, A. J. (1999). Comparison of an ordinal and a continuous outcome measure of muscle soreness. *Int J Technol Assess Health Care*, **15** (4), 709-716.

VINCENT, W. J. and WEIR, J. P. (2012). *Statistics in Kinesiology*. (4th Edition) United State: Human Kinetics.

WALKER-BONE, K., BYNG, P., LINAKER, C., READING, I., COGGON, D., PALMER, K., et al. (2002). Reliability of the Southampton examination schedule for the diagnosis of upper limb disorders in the general population**.** *Ann. Rheum Dis***, 61**, 1103-1106.

WALKER-BONE, K., PALMER, K., READING, I., COOPER, C. (2003). Soft-tissue rheumatic disorders of the neck and upper limb: prevalence and risk factors. *Seminars in arthritis rheumatism*, **33**, 185-203.

WALKER-BONE, K., READING, I., COGGON, D., COOPER, C., PALMER, K. (2004). The anatomical pattern and determinants of pain in the neck and upper limbs: epidemiologic study. *Pain*, **109**, 45-51.

WALKER-BONE, K., PALMER, K., READING, I., COGGON, D., COOPER, C. (2004). Prevalence and impact of musculoskeletal disorders in the upper limb in the general population. *Arthritis Rheum*, **51**, 642-651.

WEBB, R., BRAMMAH, T., LUNT, M., URWIN, M., ALLISON, T., SYMMONS, D. (2003). Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. *Spine,* **28**, 1195-1202.

WESTPHAL, T., PIATEK, S., SCHUBERT, S., et al. (2002). [Reliability and validity of the upper limb DASH questionnaire in patients with distal radius fractures] Reliabilitat und Validitat des Fragebogens “Upper-Limb-DASH” bei Patienten mit distalen Radiusfrakturen. *Z Orthop Ihre Grenzgeb,* **140**, 447–51.

WEWERS, M. E. and LOWE, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health*, **13** (4), 227-236.

WHO. (2001). *International Classification of Functioning, Disability and Health*. ICF.

YLINEN, J., HAKKINEN, A., et al. (2007). Neck muscle training in the treatment of chronic neck pain: a three-year follow-up study. *Eura Medicophys*, **43** (2), 161-169.

YOUNG, N. L., WILLIAMS, J. I., YOSHIDA, K. K., BOMBARDIER, C., WRIGHT, J. (1996). The context of measuring disability: Does it matter whether capability or performance is measured? *J Clin Epidemiol*, **49**, 1097–1101.

Appendix 1

Data Extraction Form - Systematic Review

PUBLICATION DETAILS

- ***Author***:
- ***Title***:
- ***Journal/source:***
- ***Origin***:

Note:
……
……

STUDY POPULATION & SAMPLING PROCEDURES

Was the sample in which the measure was validated adequately described?

1- Study (sample) size?	
2- Median or mean age? (With SD or range)	
3- Distribution of gender?	
4- Setting(s) in which the study was conducted (e.g. general population, primary care, hospital, rehab centred)?	
5- Countries in which the study was conducted?	
6- Language in which the measure was evaluated?	
7- was the method used to select patients adequately described? (e.g. convenience, constructed, random?)	
8- Was the percentage of missing responses (response rate) acceptable? (Report rate)	
9- Important disease characteristics (e.g. diagnosis, severity, status, duration etc.)	
10- Description of treatment?	
11- Additional information of relevance to study population?	

CHARACTERISTIC OF MEASURES INCLUDED IN THE STUDY

Table 1 ***- Please list all measure*** included in the study (number of measures will link to table 2)

	PROM-specific Information					
Name of the measure **(Original / modification? - please detail M in text)**	Original ref?	Contact details?	Repro?	Self-completion?	Mode of completion	Type
1-						
2-						
3-						
4-						
5-						
6-						
7-						
8-						
9-						
10-						

Footnote:

Name of the measure: please enter the name as reported by the study author.
Original ref: does the author provide the original reference for the measure?

Contact details: does the author provide contact details for the developer?
Repro: is sufficient information provided to support reproduction?
How was the measure completed?
Mode of completion?
Type of measure?

Response options:

Full name and acronym.
Please indicate if original version (**O**) or modification (**M**). if modification – please detail in item 2
Yes (√), No (x), or Not clear? If Yes = please note author and year
Yes (√), No (x), or Not clear?
Pt=patient; C=clinician; int=interview; Px=proxy (detail); O=other (detail); NC=not clear
Pen and paper; Computer; Web-based; Other
G- generic; HU-health utility; CS-condition-specific; DS-domain-specific; PS-population specific

Table 2 - ***which measurement*** and/or practical properties are evaluated in this study?

	O	**Reliability**			**Validity**						**(I) Responsiveness**
Measure (Table 3.1)		(A) Internal consistency	(B) Test-retest Reliability Inter/intra-rater reliability	(C) Measurement error	(D) Content validity	Construct validity			(H) Criterion validity		
						(E) Structural	(F) Hypothesis Testing	(G) Cross-cultural Validity	Concurrent	Predictive	
1-											
2-											
3-											
4-											
5-											
6-											
7-											
8-											
9-											
10-											

O. Does the study describe original development of the measure? Yes √ No x

Reliability – Does the study report evidence of reliability? Yes √ No x

(A) Internal Consistency
(B) Test-retest / Inter / Intra-rater reliability
(C) Evidence of measurement error

Validity – Does the study report evidence of validity? Yes √ No x

(D) Content validity
(E) Construct validity – Structural
(F) Construct validity – Hypothesis testing

- Convergent or divergent
- Known group

(G) Construct validity - Cross-cultural validity

(H) Criterion validity

- Concurrent validity
- Predictive validity

Responsiveness – Does the study report evidence of responsiveness? Yes √ No x

(I) Responsiveness

Practical properties:

Evidence of practical properties such as precision, acceptability, and feasibility will be acquired or deduced for all relevant outcome measures.

Precision – Does the study provide any evidence of data quality for the measure? (e.g. end effects, missing value etc.) Yes √ No x

Acceptability – Does the study report evidence of acceptability? (e.g. completion rate, completion time, missing value at item level etc.) Yes √ No x

Feasibility – does the study report evidence of feasibility? (e.g. time taken to complete/administer the measure reported, cost of using the measure etc.) Yes √ No x

3.3 Additional information relevant to measures listed above (e.g. modification, contact details etc.):

……

3.4 Other factors relevant to measurement application:

……

MEASUREMENT AND PRACTICAL PROPERTIES

1. RELIABILITY

1.1 Internal consistency

Is there evidence of internal consistency reliability?
Yes – Please complete table 1.1.1 (A) and 1.1.2
No
Not clear
Not applicable

Is there evidence of Item-total correlation?
Yes
No
Not clear
Not applicable

TABLE 1.1.1 - (A) COSMIN Checklist INTERNAL CONSISTENCY

Box A. Internal consistency				
1- Does the scale consist of effect indicators, i.e. is it based on reflective model?	Excellent	Good	Fair	Poor
2- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
3- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
4- Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
5- Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis not performed and no reference to another study

6- Was **the** sample size included in the unidimensionality analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 6-7* #items but <100	5* #items but <100	<5* #items
7- Was an internal consistency statistic calculated for each (unidimensional) (sub) scale separately?	Internal consistency statistics calculated for each subscale separately			Internal consistency statistics not calculated for each subscale separately
8- Were there any important flaws in the design or methods of the study?	No other important methodological flaws In the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 9- for classical Test Theory (CTT), continuous scores: was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlation calculated
10- for CTT, dichotomous scores: was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No cronbach's alpha or KR-20 and no item-total correlations calculated
11-for IRT: was a goodness of fit statistics at a global level calculated? E.g. X2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistics at a global level calculated.			Goodness of fit statistics at a global level NOT calculated.
TOTAL SCORE: ……				

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

TABLE 1.1.2 Evidence of INTERNAL CONSISTENCY - RESULTS

Name of the measure: Index Score and/or Domains	Internal Consistency Reliability			
	Population (n)	Unidimensionality of scale confirmed? (i.e in this study or by reference to other study)	Statistical analysis (e.g. alpha) and results	Item-total Correlation

Other comments specific to reliability:

……

Table 1.1.2 Guide for data extraction: Internal Consistency reliability

Index score and/or Domains: please list name of measure / index score and/or separate domains of which evidence is reported
Population and size: is evidence reported in the study population or other?
Adequacy of sample size
SP - Study Population
O - Other please summarize (plus n=)

Statistical analysis and result:

For Classic Test Theory (CTT) – Was Cronbach's alpha calculated? Indicate if different analysis. Report statistical value (and confidence intervals if reported).
For dichotomous scores – Was Cronbach's alpha or KR-20 calculated? Indicate if different analysis. Report statistical value (and confidence interval if reported).
For Item Response Theory (IRT) - Was goodness of fit statistic at a global level calculated? E.g. x2, reliability Coefficient of estimated latent trait value (index of (subject or item (separation).
Item-total correlation: please report statistical value. If reported for each item.
Please report the range of values. If values less than 0.4 are reported highlight these specific items

1.2 Test-retest / Intra-tester / Inter-tester reliability

Is there evidence of test-retest / Intra / Inter-tester reliability?
Yes – pleas complete table 1.2.1 (B) and 1.2.2
No
Not clear
Not applicable

TABLE 1.2.1 - (B) COSMIN Checklist RELIABILITY

Box (B) Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)				
Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- Were at least two measurements available?	At least two measurements			Only one measurement

	available			
5- Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	Measurements **NOT** independent
6- Was the time interval stated?	Time interval stated		Time interval not stated	
7- Was patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8- Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9- Were the test condition similar for both measurement? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10- Were any important flaws in the design or methods of the study?	No other methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 11- for continuous scores: was an interclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated and model or formula not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or with evidence that systematic change has occurred	No ICC or Pearson or Spearman correlation calculated
12- for dichotomous/nominal/ordinal scores: was Kappa calculated?	Kappa calculated			Only percentage

				agreement calculated
13- for ordinal scores: was a weighted Kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14- for ordinal scores: was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		
TOTAL SCORE: ……				

TABLE 1.2.2 Evidence of RELIABILITY - RESULTS

Name of measure: **Index score and/or domains**	**Test-retest / inter/intra-tester reliability**			
	Population (n)	Number of measurement/independence of administration/retest period/raters	Stability / similarity of test conditions	Statistical analysis and result

1.3 Measurement error (absolute measures)

Is there evidence of measurement error?

Yes – Name of the measure: ……

Statistical analysis and result: ……

Evidence of Smallest Detectable Difference: ……

Evidence of Minimal Important Change: ……

No
Not clear
Not applicable

Other comments specific to measurement error:
……
……

TABLE 1.3 - (C) COSMIN Checklist MEASUREMENT ERROR

Box (C) Measurement error: absolute measures				
Design requirement	Excellent	Good	Fair	Poor
1- was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2- was there a description of how missing items were handled?	Described how missing items were handled	NOT described it can be deduced how missing items were handled	NOT clear how missing items were handled	
3- as the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- were at least two measurements available?	At least two measurements			Only one measurement
5- were the administrations independent?	Independent measurements	Assumable that the measurement were independent	Doubtful whether the measurements were independent	Measurements NOT independent
6- was the time interval stated?	Time interval stated		Time interval NOT stated	
7- were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable patients were stable	Unclear if patients were stable	Patients were not stable
8- Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9- were the least conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or the execution of the study		Other minor methodological flaws in the design or	Other important methodological flews in the design or execution of the study

			execution of the study	
Statistical methods 11- for CTT: was the standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population
TOTAL SCORE: ……				

2. VALIDITY

2.1 Content validity (face validity)

Is there evidence of the evaluation of content or face validity? (Please tick one box)
Yes – please complete Table 2.1.1 (D)
No
Not clear
Not applicable

2.1.1 **Name of the measure:** .....

a. Development paper: YES / NO

- Measurement aim clear: YES / NO
- Conceptual basis / construct clear: YES / NO
- Purpose of the measure defined: YES / NO
- Target population defined: YES / NO
- Qualitative evidence from pre-testing with 'experts' - comprehensiveness confirmed: YES / NO

b. Application of measure in population for which it was not originally developed: YES / NO

- Measurement aim clear: YES / NO
- Conceptual basis / construct clear: YES / NO
- Purpose of the measure defined: YES / NO
- Target population defined: YES / NO
- Qualitative evidence from pre-testing with 'experts' - comprehensiveness confirmed: YES / NO

c. Application of measure for PURPOSE for which it was NOT originally developed: YES / NO

- Measurement aim clear: YES / NO

- Conceptual basis / construct clear: YES / NO
- Purpose of the measure defined: YES / NO
- Target population defined: YES / NO
- Qualitative evidence from pre-testing with 'experts' - comprehensiveness confirmed: YES / NO

TABLE 2.1.1 - (D) COSMIN Checklist - CONTENT VALIDITY

Box (D) Content Validity (Including face validity)				
General requirement	Excellent	Good	Fair	Poor
1- was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured
2- was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3- was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (Discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4- was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehensively reflect the construct to be measured
5- were there any important flaws in the design or methods of the study	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
TOTAL SCORE: ……				

2.2 CONSTRUCT validity

2.2.1 STRUCTURAL validity

Is there evidence of structural (Internal construct) validity? E.g. principle component analysis; factor analysis
Yes – please complete Table 2.2.1 (E) and results
No
Not clear

TABLE 2.2.1 - (E) COSMIN Checklist STRUCTURAL validity

Box (E) Structural validity				
1- does the scale consists of effect indicators, i.e. it based on a reflective model?	Excellent	Good	Fair	Poor
Design requirement 2- was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3- was there a description of how missing items were handled?	Described how missing items were handled	Not described how missing items were handled	Not clear how missing items were handled	
4- was the sample size included in the analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 5-7* #items but <100	5* #items but <100	<5* #items
5- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation methods not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)
Statistical methods 6- for CTT: was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	Exploratory factor analysis performed while confirmatory would have been more appropriate		No exploratory or confirmatory factor analysis performed
7- for IRT: were IRT tests for determining the (uni-) dimensionality of the items performed?	IRT test for determining (uni) dimensionality performed			IRT test determining (uni) dimensionality NOT Performed
TOTAL SCORE: ……				

Structural validity – Results

a. Name of measure:……
Statistical analysis and result (include population and n=):……
……
……
……
……

b. Name of measure:……
Statistical analysis and result (include population and n=):……
……
……
……

2.2.2 CROSS-CULTURAL validity (Translated Questionnaire Only)

Table 2.2.2 - (G) COSMIN checklist CROSS-CULTURAL Validity

Box (G) Cross-cultural validity				
Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	CTT: 7* #items and ≥100 IRT: ≥200 per group	CTT: 5* #items and ≥100 OR 5-7* #items but <100 IRT: ≥200 in 1 group and 100- 199 in 1 group	CTT: 5* #items but <100 IRT: 100-199 per group	CTT: <5* #items IRT: (<100 in 1 or both groups
4- where both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language not known

5 Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the construct to be measured, expertise in both languages	Expertise of the translators described with respect to disease, construct, and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6- Did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independent
7- Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translation but one backward translation	One forward and one backward translation	Only a forward translation
8- Was there an adequate description of how differences between the original and translated version were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		
9. Was the translation reviewed by committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		
10. Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested, in the	Translated instrument pre-tested, but unclear if this was done in the target population	Translated instrument pre-tested, but NOT	Translated instrument NOT pre-tested

	target population		in the target population	
11. Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		Sample used in the pre- test NOT (adequately) described	
12. Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language/culture	Stated (but not shown) that samples were similar for all characteristics except language/culture	Unclear whether samples were similar for all characteristics except language/culture	Samples were NOT similar for all characteristics except language/culture
13. Were there any important flaws in the design or methods of the study?	No other methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 14- for CTT: was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed			Multiple-group confirmatory factor analysis NOT performed
15- for IRT: was differential item function (DIF) between language groups assessed?	DIF between language groups assessed			DIF between language groups NOT assessed.
TOTAL SCORE: ……				

2.2.3 HYPOTHESES TESTING COSMIN Checklist (F)

Is there evidence of construct (Convergent / divergent) validity? Please tick one box

Yes – please complete Table 2.2.3.1 (F) and 2.2.3.2
No
Not clear

Is there evidence of known groups validity? Please tick one box

Yes – please complete Table 2.2.3.1 (F) and 2.2.3.2
No
Not clear

Was a hypothesis to be tested stated priori?
Yes – please complete Table 2.2.3.1 (F) and 2.2.3.2
No
Not clear

TABLE 2.2.3.1 - (F) COSMIN checklist - HYPOTHESIS TESTING

Box (F) Hypothesis Testing				
Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- Were hypotheses regarding correlation or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5- was the expected direction of correlation or mean differences included in the hypotheses?	Expected direction of the correlation or differences	Expected direction of the correlation or differences NOT stated		

	stated			
6- Was the expected absolute or relative magnitude of correlation or mean differences included in the hypotheses?	Expected magnitude of the correlation or differences stated	Expected magnitude of the correlation or differences NOT stated		
7- for convergent validity: was an adequate description provided of the comparator instrument(s)?	Adequate description of the construct measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	No description of the constructs measured by the comparator instrument(s)
8- for convergent validity: were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
9- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measure another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods 10- were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlation applied, but distribution of the score	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

		or mean (SD) not presented		
TOTAL SCORE: ……				

TABLE 2.2.3.2a: evidence of CONSTRUCT VALIDITY: Name of the measure: ……

Name of the test / measure / known groups:		**Hypothesised relationship**		
	Domain measured	**Stated priori?**	**Confirmed?**	**Result**

TABLE 2.2.3.2b: evidence of CONSTRUCT VALIDITY: name of measure: ……

Name of the test / measure / known groups:		**Hypothesised relationship**		
	Domain measured	**Stated priori?**	**Confirmed?**	**Result**

TABLE 2.2.3.2c: evidence of CONSTRUCT VALIDITY: name of measure: ……

Name of the test / measure / known groups:		**Hypothesised relationship**		
	Domain measured	**Stated priori?**	**Confirmed?**	**Result**

Table 2.2.3.2 a, b, and c: Guide for data extraction

Name of test / measure: Please list

Domain: Please list

Was a hypothesised relationship between measures / domains proposed a priori?

YES
NO
NOT CLEAR

Was a hypothesised relationship between measures / domains confirmed?
YES
NO
NOT CLEAR

Result of correlation: please report: ……

2.3 CRITERION validity (Concurrent and Predictive)

TABLE 2.3 - (H) COSMIN checklist CRITERION VALIDITY

Box (H) Criterion validity including concurrent and predictive				
Design requirements	Excellent	Good	Fair	Poor
1. Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2. Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3. Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4. Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
5. Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

	study			
Statistical methods 6. For continuous scores: were correlation, or the area under the receiver-operating curve calculated?	Correlation or AUC calculated			Correlation or AUC NOT calculated
7. For dichotomous scores: were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated
TOTAL SCORE: ……				

3. RESPONSIVENESS

Is there evidence of responsiveness?
YES (Please complete table 3.1 (I) and 3.2
NO
NOT CLEAR

Was a hypothesis to be tested stated a priori?
YES (please briefly state)
NO
NOT CLEAR

TABLE 3.1 - (I) COSMIN Checklist RESPONSIVENESS

Box (I) Responsiveness				
Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		

2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- was the longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
5- was the time interval stated?	Time interval adequately described			Time interval NOT described
6- if anything occurred in the interim period (e.g. intervention, other relevant event), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period	
7-was the proportion of the patients changed (i.e. improvement or deterioration) ?	Part of the patients were changed (evidence provided)	No evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed	Patients were NOT changed
Design requirement for hypotheses testing For construct for which a gold standard was not available: 8- were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
9- was the expected direction or correlation or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlation or differences NOT stated		
10- Were the expected absolute or relative magnitude of correlation or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
11- was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the construct measured by the comparator instrument(s)	No description of the constructs measured by the comparator instrument(s)

12- Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
13- were any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measure another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods 14- were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate
Design requirement for comparison to a gold standard For construct for which a gold standard was available: 15- can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
16- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 17- for continuous scores: were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated

18- for dichotomous scales: were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

TABLE 3.2: evidence of RESPONSIVENESS **(please complete table for each measure with evidence of responsiveness. (Additional details in following text)**

Name of the measure	Condition/Intervention/Criterion for change	N	Follow-up	Baseline mean (SD)	Follow-up mean (SD)	Mean change in score (p value)	Reported evidence of responsiveness		
							Distribution-based (ES)	Anchor-based (external criterion: p value)	Correlation of change score (p value)

Guide for data extraction

Name of the measure:	
Condition / intervention / criterion for change	Type of intervention and study (e.g.RCT) / health transition question etc.
Period of follow-up	As reported by author
Type of evidence reported	Please inter results OR indicate if not reported (NR) or not clear (NC) etc.

RESPONSIVENESS Where appropriate please complete detail for each measure:

3.2.1 **Detail re intervention and/or Criterion change in health:**

- **Intervention (known efficacy?): ……**
- **Criterion for change in health: ……**
- **Health transition questions: ……**
- **Hypothesised association between intervention and outcome stated a priori:** ……

3.2.2 **Evidence of distribution-based assessment:**

NOT CLEAR

NO

YES (please detail below and complete Table 3.2): duration of follow-up: ……

Effect size: Value: ……
Standardised response mean: Value: ……
Modified standardised response mean: Value: ……
Other: Value: ……

3.2.3 **Evidence of Anchor-based assessment:**
NOT CLEAR
NO
YES (please detail below and complete table 3.2): duration of follow-up: ……

- **External anchor – e.g. external measure of change in health: ……**
- **Statistical analysis and result:** ……

3.2.4 **Correlation of change scores**
NOT CLEAR
NO
YES ((please detail below and complete table 3.2): duration follow-up: ……

- **Correlation between change scores in which measures: ……**
- **Statistical analysis and result:** ……

3.2.5 **Does the study report mean change in score?**
NOT CLEAR
NO
YES (please detail below and complete table 3.2): duration of follow-up: ……

- **Statistical analysis and results:** ……

3.2.6 **Is other evidence of measurement responsiveness reported? For example ROC analysis**
NOT CLEAR
NO
YES (please detail below and complete table 3.2): duration of follow-up: ……

- **Statistical analysis and results:** ……

3.2.7 **Other comments specific to responsiveness: ……**
……

3. PRACTICAL PROPERTIES

4.1 PRECISION

4.1.1 Does the study describe measurement end effects?
Not reported
Yes (please give detail)

Name of the measure:
No evidence of end effects
Floor effects: % floor: ……
Ceiling effects: % ceiling: ……

Other comment specific to precision: ……

4.2 ACCEPTABILITY

Name of the measure(s): ……

4.2.1 Are measurement completion rates (response rate) reported?
NO
NOT CLEAR
YES (please detail): **……**

4.2.2 Are missing values reported at item level (i.e. items omitted more frequently than other items):
No
NOT CLEAR
YES (please detail): ……

4.2.3 Is completion time reported?
NO
NOT CLEAR
YES (please detail): ……

4.2.4 Is the reading / comprehension level reported?
NO
NOT CLEAR
YES (please detail): **……**

4.2.5 Are any special requirements placed on respondents?
NO
NOT CLEAR
YES (please detail): **……**

4.2.6 Were the views of patients explicitly explored with regard to the measure?
NO
NOT CLEAR
YES (please detail – include population details if different from main study): **……**

4.2.7 Other comments specific to acceptability (complete as necessary): ……

4.3 FEASIBILTY

Name of the measure(s): ……

4.3.1 Was the time taken to administer / complete the measure reported?
NO
NOT CLEAR
YES (please detail): **……**

4.3.2 Was the time taken to score the measure reported?
NO
NOT CLEAR
YES (please detail): **……**

4.3.3 **Is the cost of using the measure reported? For example, purchasing the license?**
NO
NOT CLEAR
YES (please detail): **……**

4.3.4 Is there a need for technological or instruction support when using the measure?
NO
NOT CLEAR
YES (please detail): **……**

4.3.5 Is there a need for staff training to support application of the measure?
NO
NOT CLEAR
YES (please detail): **……**

4.3.6 Other comments specific to feasibility: ……

5. INTERPRETABILITY (J)

Interpretability box is used to extract all information on the interpretability issues described in this box of the instrument under study from included articles. **INTERPRETABILITY IS NOT A MEASUREMENT PROPERTY TO BE RATED/ASSESSED**.

TABLE 5 (J) (COSMIN Checklist) – INTERPRETABILITY

Box (J) Interpretability	
1. Percentage of missing items	
2. Description of how missing items were handled	
3. Distribution of the (total) score	
4. Percentage of the respondents who had the lowest possible (total) score	
5. Percentage of respondents who had the highest (total) score	
6. Scores and change scores (i.e. mean and SD) for relevant (sub) groups e.g. for normative groups, subgroups of patients, or general population	
7. Minimal important change (MIC) or minimal important differences (MID)	

5.1 Interpretability – the author reports evidence in support of: (please tick all that apply and provide detail if possible)

5.1.1 Name of the measure: ……

- Minimal important change (MIC): ……
- Minimal clinically important change (MCIC): ……

- Smallest detectable change (SDC): ……
- Patient acceptable symptom state (PASS): ……
- Other: please describe: ……

5.1.2 Name of the measure:

- Minimal important change (MIC): ……
- Minimal clinically important Change (MCIC): ……
- Smallest detectable Change (SDC): ……
- Patient acceptable symptom state (PASS): ……
- Other: please describe: ……

5.1.3 Other comments specific to interpretability: (population groups / external criterion / intervention of known efficacy etc

……
……

6. ADDITIONAL COMMENTS

Please list any references from the article, which should be obtained for future review (please list ref number, author, and journal details):
……
……

Appendix 2

COSMIN Checklist with 4-Point Scale

Contact
CB Terwee, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
1081 BT Amsterdam
The Netherlands
Website: www.cosmin.nl, www.emgo.nl
E-mail: cb.terwee@vumc.nl

Instructions
This version of the COSMIN checklist is recommended for use in systematic reviews of measurement properties. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. A methodological quality score per box is obtained by taking the lowest rating of any item in a box ('worse score counts'). For example, if for a reliability study one item in the box 'Reliability' is scored poor, the methodological quality of that reliability study is rated as poor. The Interpretability box and the Generalizability box are mainly used as data extraction forms. We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box (e.g. norm scores, floor-ceiling effects, minimal important change) of the instruments under study from the included articles. Similar, we recommend to use the Generalizability box to extract data on the characteristics of the study population and sampling procedure. Therefore, no scoring system was developed for these boxes.

This scoring system is described in this paper:

Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research 2011, July 6.

Step 1. Evaluated measurement properties in the article

	Internal Consistency	Box A
	Reliability	Box B
	Measurement Error	Box C
	Content Validity	Box D
	Structural Validity	Box E
	Hypothesis Testing	Box F
	Cross-cultural Validity	Box G
	Criterion Validity	Box H
	Responsiveness	Box I

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

	Excellent	Good	Fair	Poor
1- Was the IRT model used adequately described? e.g. one Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	IRT model adequately described	IRT model not adequately described		
2- Was the computer software package used adequately described? e.g. RUMM 2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	Software package adequately described	Software package not adequately described		
3- Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	Method of estimation adequately described	Method of estimation not adequately described		
4- Were the assumptions for estimating parameters of the IRT model checked? E.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	Assumption of the IRT model checked	Assumption of the IRT model party checked	Assumption of the IRT model not checked or unknown	

To obtain a total score for the methodological quality of studies that use IRT methods, the 'worse score count' algorithm should be applied to the IRT box in combination with the box of the measurement property that was evaluated in the IRT study. For example, if IRT methods are used to study internal consistency and item 4 in the IRT box is scored fair, while the items in the internal consistency box (box A) are all scored as good or excellent, the methodological quality score for internal consistency will be fair. However, if any of the items in box A is scored poor, the methodological quality score for internal consistency will be poor.

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency

	Excellent	Good	Fair	Poor
1- Does the scale consist of effect indicators, i.e. is it based on reflective model?				
2- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
3- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
4- Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
5- Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was	Authors refer to another study in which factor	Factor analysis not performed

		performed in a similar study population	analysis was performed, but not in a similar study population	and no reference to another study
6- Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 6-7* #items but <100	5* #items but <100	<5* #items
7- Was an internal consistency statistic calculated for each (unidimensional) (sub) scale separately?	Internal consistency statistics calculated for each subscale separately			Internal consistency statistics not calculated for each subscale separately
8- Were there any important flaws in the design or methods of the study?	No other important methodological flaws In the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 9- for classical Test Theory (CTT), continuous scores: was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlation calculated
10- for CTT, dichotomous scores: was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No cronbach's alpha or KR-20 and no item-total correlations calculated
11-for IRT: was a goodness of fit statistics at a global level calculated? E.g. X2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistics at a global level calculated.			Goodness of fit statistics at a global level NOT calculated.

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest, inter- and intra-rater reliability).

Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- Were at least two measurements available?	At least two measurements available			Only one measurement

5- Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	Measurements **NOT** independent
6- Was the time interval stated?	Time interval stated		Time interval not stated	
7- Was patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8- Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9- Were the test condition similar for both measurement? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10- Were any important flaws in the design or methods of the study?	No other methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 11- for continuous scores: was an interclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated and model or formula not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or with evidence that systematic change has occurred	No ICC or Pearson or Spearman correlation calculated
12- for dichotomous/nominal/ordinal scores: was Kappa calculated?	Kappa calculated			Only percentage agreement calculated
13- for ordinal scores: was a weighted Kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14- for ordinal scores: was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures

Design requirement	Excellent	Good	Fair	Poor
1- was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2- was there a description of how missing items were handled?	Described how missing items were handled	NOT described it can be deduced how missing items were handled	NOT clear how missing items were handled	

3- as the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- were at least two measurements available?	At least two measurements			Only one measurement
5- were the administrations independent?	Independent measurements	Assumable that the measurement were independent	Doubtful whether the measurements were independent	Measurements NOT independent
6- was the time interval stated?	Time interval stated		Time interval NOT stated	
7- were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable patients were stable	Unclear if patients were stable	Patients were not stable
8- Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9- were the least conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or the execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flews in the design or execution of the study
Statistical methods 11- for CTT: was the standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population

Box D. Content validity (including face validity)

General requirement	Excellent	Good	Fair	Poor
1- was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured

2- was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3- was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (Discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4- was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehensiv ely reflect the construct to be measured
5- were there any important flaws in the design or methods of the study	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodologica l flaws in the design or execution of the study

Box E. Structural validity

1- does the scale consists of effect indicators, i.e. it based on a reflective model?	Excellent	Good	Fair	Poor
Design requirement 2- was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3- was there a description of how missing items were handled?	Described how missing items were handled	Not described how missing items were handled	Not clear how missing items were handled	
4- was the sample size included in the analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 5-7* #items but <100	5* #items but <100	<5* #items
5- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodologica l flaws in the design or execution of the study (e.g. rotation methods not described)	Other important methodologi cal flaws in the design or execution of the study (e.g. inappropriat e rotation method)
Statistical methods 6- for CTT: was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	Exploratory factor analysis performed while confirmatory would have been more appropriate		No exploratory or confirmatory factor analysis performed

7- for IRT: were IRT tests for determining the (uni-) dimensionality of the items performed?	IRT test for determining (uni) dimensionality performed			IRT test determining (uni) dimensionality NOT Performed

Box F. Hypotheses testing

Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- Were hypotheses regarding correlation or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5- was the expected direction of correlation or mean differences included in the hypotheses?	Expected direction of the correlation or differences stated	Expected direction of the correlation or differences NOT stated		
6- Was the expected absolute or relative magnitude of correlation or mean differences included in the hypotheses?	Expected magnitude of the correlation or differences stated	Expected magnitude of the correlation or differences NOT stated		
7- for convergent validity: was an adequate description provided of the comparator instrument(s)?	Adequate description of the construct measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	No description of the constructs measured by the comparator instrument (s)
8- for convergent validity: were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument (s)
9- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only	Other important methodological flaws in the design or

			data presented on a comparison with an instrument that measure another construct)	execution of the study
Statistical methods 10- were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlation applied, but distribution of the score or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Box G. Cross-cultural validity

Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	CTT: 7* #items and ≥100 IRT: ≥200 per group	CTT: 5* #items and ≥100 OR 5-7* #items but <100 IRT: ≥200 in 1 group and 100- 199 in 1 group	CTT: 5* #items but <100 IRT: 100-199 per group	CTT: <5* #items IRT: (<100 in 1 or both groups
4- where both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language not known
5- was the expertise of the people involved in the translation process adequately described? e.g. expertise in the construct to be measured, expertise in both languages	Expertise of the translators described with respect to disease, construct, and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6- did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independe nt
7- Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translation but one backward translation	One forward and one backward translation	Only a forward translation
8- was there an adequate description of how differences between the original and translated version were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		

9- Was the translation reviewed by committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		
10- Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested, in the target population	Translated instrument pre-tested, but unclear if this was done in the target population	Translated instrument pre-tested, but NOT in the target population	Translated instrument NOT pre-tested
11- Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		Sample used in the pre-test NOT (adequately) described	
12- Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language/culture	Stated (but not shown) that samples were similar for all characteristics except language/culture	Unclear whether samples were similar for all characteristics except language/culture	Samples were NOT similar for all characteristics except language/culture
13- Were there any important flaws in the design or methods of the study?	No other methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 14- for CTT: was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed			Multiple-group confirmatory factor analysis NOT performed
15- for IRT: was differential item function (DIF) between language groups assessed?	DIF between language groups assessed			DIF between language groups NOT assessed.

Box H. Criterion validity

Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample	Small sample size (<30)

			size (30-49)	
4- can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
5- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 6- for continuous scores: were correlation, or the area under the receiver-operating curve calculated?	Correlation or AUC calculated			Correlation or AUC NOT calculated
7- for dichotomous scores: were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Box I. Responsiveness

Design requirements	Excellent	Good	Fair	Poor
1- Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items not described		
2- Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
3- Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4- was the longitudinal design with at least two measurements used?	Longitudinal design used			No longitudinal design used
5- was the time interval stated?	Time interval adequately described			Time interval NOT described
6- if anything occurred in the interim period (e.g. intervention, other relevant event), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the	

			interim period	
7-was the proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	No evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed	Patients were NOT changed
Design requirement for hypotheses testing For construct for which a gold standard was not available: 8- were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
9- was the expected direction or correlation or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlation or differences NOT stated		
10- Were the expected absolute or relative magnitude of correlation or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
11- was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the construct measured by the comparator instrument (s)	No description of the constructs measured by the comparator instrument (s)
12- Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument (s) in any study population	No information on the measurement properties of the comparator instrument (s)
13- were any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution	Other important methodological flaws in the design or execution

			of the study (e.g. only data presented on a comparison with an instrument that measure another construct)	of the study
Statistical methods 14- were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate
Design requirement for comparison to a gold standard For construct for which a gold standard was available: 15- can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
16- were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods 17- for continuous scores: were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated
18- for dichotomous scales: were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Interpretability

We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box of the instruments under study from the included articles.

1- Percentage of missing items	
2- Description of how missing items were handled	
3- Distribution of the (total) score	

4- Percentage of the respondents who had the lowest possible (total) score	
5- Percentage of the respondents who had the highest possible (total) score	
6- Scores and change scores (i.e. mean and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or general population	
7- Minimal important change (MIC) or minimal important differences (MID)	

Generalizability

We recommend to use the Generalizability box to extract data on the characteristics of the study populations and sampling procedures of the included studies.

1- Median or mean age (with standard deviation or range)	
2- Distribution of sex	
3- Important disease characteristics (e.g. severity, status, duration) and description of treatment	
4- Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)	
5- Countries in which the study was conducted	
6- Language in which the HR-PRO instrument was evaluated	
7- Methods used to select patients (e.g. convenience, consecutive, or random)	
8- Percentage of missing responses (response rate)	

Appendix 3

Summary of Quality of the Measurement Properties (Systematic Review)

Measurement properties that were tested in each study are reported below.
Items with the worst rating only are reported for each measurement property.
Yellow highlight indicates the rating assigned to each measurement property.

Study 1:
Author: Huisstede et al. (2009)
Title: Is the Disability of Arm, Shoulder, and Hand Questionnaire (DASH) Also Valid and Responsive in patients with Neck Complaints.
Journal/Source: Spine Volume 34, pp E130 – E138
Origin: Netherland
Tested Properties: Hypothesis testing and Responsiveness

Table 1: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Construct Validity (Hypothesis Testing) Item 8	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Responsiveness Item 12	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Item 13	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measure another construct)	Other important methodological flaws in the design or execution of the study
Item 14	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Study 2:
Author: Mehta et al. (2010)
Title: Concurrent validation of the DASH and the QDASH in comparison to neck-specific scales in patients with neck pain
Journal/source: Spine
Origin: Canada
Tested Properties: Hypothesis testing and Criterion validity

Table 2: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Construct Validity (Hypothesis Testing) Item 8	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Criterion Validity Item 3	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)

Study 3:
Author: Fan et al. (2008)
Title: Assessing validity of the QuickDASH and SF-12 as surveillance tools among workers with neck or upper extremity musculoskeletal disorders
Journal/source: Journal Hand Therapy
Origin: USA
Tested Properties: Hypothesis testing and Criterion validity

Table 3: Quality Rating

Measurement Property	**Excellent**	**Good**	**Fair**	**Poor**
Construct Vallidity (Hypothesis Testing) Item 4	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
Item 9	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measure another construct)	Other important methodological flaws in the design or execution of the study
Item 10	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlation applied, but distribution of the score or mean	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

		(SD) not presented		
Criterion Validity Item 4	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'

Study 4:
Author: Fan et al. (2011)
Title: Responsiveness of the QuickDASH and SF-12 in workers with neck or upper extremity musculoskeletal disorders: one-year follow-up.
Journal/source: J Occup Rehabil
Origin: USA
Tested Properties: Responsiveness

Table 4: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Responsiveness 1tem 14	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Study 5:
Author: Stock et al., 2003
Title: The impact of neck and upper limb musculoskeletal disorders on the lives of affected workers: development of a new functional status index.
Journal/source: Qual Life Res
Origin: Canada
Tested Properties: Reliability (Internal consistency), (test-retest, inter-rater), content validity and construct validity (structural, hypothesis testing).

Table 5: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Reliability: (internal consistency) Item 3	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
Reliability: Test-retest/interrater Item 2	Described how missing items were handled	Not described but it can be deduced how missing items were handled	No clear how missing items were handled	
Item 7	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
Content Validity Item 1	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of construct to be measured poorly described AND this was not	NOT assessed if all items refer to relevant aspects of the construct to be measured

			taken into consideration	
Item 2	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
Item 3	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
Item 4	Assessed if all items together comprehensively reflect the construct to be measured.		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehensively reflect the construct to be measured
Item 5	No other important methodological flaws in the design or execution of the study.		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Structural Validity Item 3	Described how missing items were handled	Not described how missing items were handled	Not clear how missing items were handled	
Hypothesis Testing Item 8	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)
Responsiveness Item 14	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Study 6:
Author: Lomond and Cote
Title: Shoulder functional assessments in persons with chronic neck/shoulder pain and healthy subjects: reliability and effects of movement repetition
Journal/source: IOS Press
Origin: Canada
Tested Properties: Test-retest/inter-rater reliability, measurement error and hypothesis testing.

Table 6: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Reliability (Test-retest) Item 3	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Item 7	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
Measurement Error Item 3	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
Item 7	Patients were stable (evidence provided)	Assumable patients were stable	Unclear if patients were stable	Patients were not stable
Hypothesis Testing Item 4	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected

Study 7:
Author: Manik Kulkarni
Title: inter-rater and intra-rater reliability for the Single Arm Military Press Test on upper limb disability in adult females without neck pain
Journal/source: Physiotherapy Journal (unpublished)
Origi*n*: UK
Tested Properties: Inter- and intra-rater reliability

Table 7: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Reliability (inter- and intra-rater) Item 7	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable

Study 8:
Author: Vivek Jain
Title: Inter-rater and intra-rater reliability of Single Arm Military Press (SAMP) in female subjects with neck pain
Journal/source: International Journal of Physiotherapy and Rehabilitation (unpublished)
Origin: UK
Tested Properties: Inter- and intra-rater reliability

Table 8: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Reliability (inter- and intra-rater) Item 7	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable

Study 9:
Author: Priya Patekar
Title: Clinical utility of Single Arm Military Press (SAMP) Test in females with neck pain

Journal/source: Physiotherapy Journal (unpublished)
Origin: UK
Tested Properties: Construct validity (hypothesis testing).

Table 9: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Hypothesis Testing Item 4	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulated a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected

Study 10:
Author: Rakhi Darne
Title: Construct validity of Single Arm Military Press (SAMP) in females with neck pain
Journal/source: International Journal of Physiotherapy and Rehabilitation (unpublished)
Origin: UK
Tested Properties: Construct validity (hypothesis testing).

Table 10: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Hypothesis Testing Item 8	Adequate measurement properties of the comparator instrument(s) in a population similar to the study	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

Study 11:
Author: Balassoubramanien Toulassidharane
Title: Construct validity of the Single Arm Military Press (SAMP) in a female non-patient population
Journal/source: International Journal of Physiotherapy and Rehabilitation (Unpublished)
Origin: UK
Tested Properties: Construct validity (hypothesis testing).

Table 11: Quality Rating

Measurement Property	Excellent	Good	Fair	Poor
Hypothesis Testing Item 8	Adequate measurement properties of the comparator instrument(s) in a population similar to the study	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on the measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

Appendix 4

SHU Ethical Approval for the Survey

Sheffield Hallam University | Centre for Health and Social Care Research

Date 17022015

Ref 2014-5/HWB/HSC/STAFF/13

Dear Ahmad Alreni

This letter relates to your research proposal: Management of neck pain - a survey of current UK physiotherapy practice.

This proposal was submitted to the Faculty Research Ethics Committee with a standard SHREC1 form. This indicates that your project is low risk and I concur with that view. As such I can approve it and have added to the register of projects and given a reference number. You do not need any further review from the Ethics Committee. You will need to ensure you have all other necessary permission in place before proceeding, for example, from the Research Governance office of any sites outside the University where your research will take place. This letter can be used as evidence that the proposal has been registered within Sheffield Hallam University.

The documents reviewed were:

ALRENI1.pdf

Good luck with your project.

Yours sincerely

Peter Allmark

Peter Allmark
Chair Faculty Research Ethics Committee
Faculty of Health and Wellbeing
Sheffield Hallam University
32 Collegiate Crescent
Sheffield
S10 2BP
0114 224 5727
p.allmark@shu.ac.uk

Appendix 5

Survey Instrument

UK National Survey of Neck Pain

Welcome

Dear Colleague

We would like to invite you to take part in this online survey on the topic: Clinical Management of Musculoskeletal Neck Pain. This survey is part of a collaborative research project investigating current physiotherapy management of neck pain within the UK.

As you know, neck pain is a common musculoskeletal condition, which has substantial socioeconomic impact on patient and society. Although various approaches are advocated for the management of patients with neck pain, it is unclear which are most commonly and least commonly used. Your response will help us understand current physiotherapy practice, which will subsequently enable us to develop an appropriate programme of research. We anticipate that the results of this survey will be published.

Please complete this survey if: 1) you are currently registered and practicing physiotherapy in the UK and 2) you have seen at least one case of neck pain in the last 6 months.

Your responses are extremely valuable to us. We would be grateful if you could take the time to complete this survey. It should take around 5-10 minutes to complete. Please be aware that your responses will be treated in strict confidence and the data will be entered anonymously onto an electronic form of data analysis. Completion of this survey is voluntary.

if you have queries regarding this survey you can email Ahmad Alreni at a.alreni@shu.ac.uk or Dr Sionnadh McLean at s.mclean@shu.ac.uk. Thank you for your time.

Your sincerely

Ahmad Alreni, PhD researcher, Centre for Health and Social Care Research, Sheffield Hallam University
Dr Karen Kilner, Statistician, Centre for Health and Social Care Research, Sheffield Hallam University

Deborah Harrop, Information Scientist, Centre for Health and Social Care Research, Sheffield Hallam University
Dr Sionnadh Mclean, Reader in Physiotherapy, Department of Allied Health Professions, Sheffield Hallam University.

Section 1

About you

This information will in no way identify you or your responses. In this survey, all responses are treated strictly confidential.

- **Are you**
  - **Male**
  - **Female**

- **Do you work**
  - Exclusively in the National Health Service (NHS)
  - Exclusively in non-NHS setting (e.g. private practice/hospital, education/research)
  - A combination of NHS and non-NHS
  - Other setting

  Please specify:

- **Years of practice**
  - Less than 2 years
  - 2-5 years
  - 6-10 years
  - 11-15 years
  - 15+ years

- **In which nation, do you practice clinically?**
  - Scotland
  - Northern Ireland
  - Wales
  - England

- **Have you completed any postgraduate training (MSc and /or PhD)?**
  - **No**
  - **Yes**

  Please specify:

- **What proportion of your caseload is made up of patients with neck pain?**
  - **Less than 25%**
  - **25-50%**
  - **51-75%**
  - **more than 75%**

- **Do you have a special interest in neck pain?**
  - **No**

- **Yes**

Further comments:

Section 2

Treatment approaches to management

In this section, we are interested in identifying the management approach/approaches you typically use for patients with non-specific neck pain. Non-specific neck pain is defined here as a dysfunction in the cervical structures NOT caused by any serious acute trauma (e.g. Whiplash Associated Disorder), systemic disease, neurological disorder (e.g. Cervical Radiculopathy, Nerve Root Compression) or inflammatory condition.

Which management approach/approaches do you use most often for patients with non-specific neck pain?

- **Therapeutic exercise**

If selected:

Which exercise component/components do you use regularly for patients with non-specific neck pain?

- ❖ General aerobic/strengthening/endurance exercise
- ❖ Cervical strengthening exercise
- ❖ Upper limb strengthening exercise
- ❖ Cervical stretching exercise
- ❖ Upper limb stretching exercise
- ❖ Cervical stabilising exercise
- ❖ Upper limb stabilising exercise
- ❖ Balance exercise
- ❖ Proprioception exercise for the eyes
- ❖ Proprioception exercise for the cervical spine
- ❖ Proprioception exercise for the upper limb
- ❖ Other

Please specify:

Manual therapy (Manipulation/Mobilisation)

If selected:

Which manual therapy method/methods do you use regularly for patients with non-specific neck pain?

- ❖ Maitland
- ❖ Mulligan
- ❖ Cyriax
- ❖ Society of Orthopaedic Medicine
- ❖ Kaltenborn
- ❖ Manipulation (Grade V)
- ❖ Other

Please specify:

- **Electrotherapy**

If selected:

Which electrotherapy method/methods do you use regularly for patients with NSNP?

- ❖ Galvanic Current (DC)
- ❖ Electrical Nerve Stimulation (ENS)
- ❖ Pulsed Electromagnetic Fields (PEMF)
- ❖ Transcutaneous Nerve Stimulation (TENS)
- ❖ Repetitive Magnetic Stimulation (rMS)
- ❖ Other

Please specify:

- **The McKenzie method (Mechanical Diagnosis and Therapy (MDT)/End-Range Exercise/Active Range of Motion Exercises (AROM)/Direction Preference Exercise/Unloaded Exercise)**
- **Therapeutic patient education (oral, written, Audio-visual, etc.)**
- **Massage therapy (all types)**
- **Acupuncture**
- **Traction**
- **Heat/cold**
- **Taping/strapping**
- **Hydrotherapy**
- **Feldenkrais**
- **Other management approach/approaches**

Please specify:

Section 3

Outcome measures

In this section, we are interested in your use of outcome measures in the assessment/management of patients with non-specific neck pain.

Do you use outcome measures in the assessment/management of patients with non-specific neck pain?

No

If selected:

This is because

- Lack of time
- Lack of clear guidance about suitability of available measures
- Lack of access to information/knowledge about outcome measures
- Lack of resources (e.g. expensive to purchase)
- There is no need to use outcome measures for patients with NSNP

- Other

Please specify:

Submit

Yes
If selected:

This is because
- Medicolegal documentation requirement
- Fulfilling charting/documentation
- Setting treatment goals
- Communicating with patients
- Communicating with other healthcare professionals
- Marketing
- Research
- Other reasons

Please specify:

Your use of outcome measures for patients with non-specific neck pain is

- **Routinely: >70% of cases**
- **Regularly: 51-70% of cases**
- **Sometimes: 11-50% of cases**
- **Rarely 1-10% of cases**

Which outcome instrument/instruments do you use most frequently for patients with non-specific neck pain?

- **Patient-reported outcome measurements (PROMs)**

(Instruments that are subject-completed relaying on patient's self-perception of pain, mobility status and performance of daily activity)

If selected:

Which PROM scale/scales do you use for patients with non-specific neck pain?

	Routinely >70% Cases	Regularly 11-70% Cases	Rarely 1-10% Cases	Never 0% Cases
Visual Analogue Scale (VAS)	○	○	○	○
Numeric Rating Scale (NRS)	○	○	○	○

Pain Distress Scale	○	○	○	○
Pain Catastrophizing Scale (PCS)	○	○	○	○
Neck Disability Index (NDI)	○	○	○	○
Whiplash Disability Questionnaire	○	○	○	○
Bournemouth Questionnaire	○	○	○	○
Patient Specific Functional Scale (PSFS)	○	○	○	○
Disability of Arm, Shoulder and Hand (DASH)/QuickDASH	○	○	○	○
Northwick Park Neck Pain (NPQ)	○	○	○	○
Neck and Upper Limb Index (NULI)	○	○	○	○
Rapid Upper Limb Assessment (RULA)	○	○	○	○
Time Lost from Work	○	○	○	○
Work Limitation Scale	○	○	○	○
Work Distress Scale	○	○	○	○
SF-36/SF12	○	○	○	○
Euro-Qol/EQ5D	○	○	○	○
WHO-Brief	○	○	○	○
Nottingham Health Profile (NHP)	○	○	○	○
Fear of Movement Scales	○	○	○	○
Depression/Anxiety Scales	○	○	○	○
Patients Global Perceived Rating of Improvement or Satisfaction	○	○	○	○

- ○ Other PROM scale

Please specify:

- **Performance–based/physical/functional measures**

(Instruments that use tasks in clinical setting to measure a patient's functional capacity)
If selected:

Which performance-based/physical/functional measure/measures do you use for patients with non-specific neck pain?

	Routinely >70% Cases	Regularly 11-70% Cases	Rarely 1-10% Cases	Never 0% Cases
Quantitative Sensory Test QST (e.g. vibrometry, touch, temperature)	○	○	○	○
Pain Algometry (e.g. pain pressure threshold tests)	○	○	○	○
Rating of Segmental Joint Mobility (Passive Accessory Motion Tests; Passive Physiological Motion Tests)	○	○	○	○
Goniometric Measures of Neck Motion	○	○	○	○
Inclinometer of Neck Motion	○	○	○	○
Movement Diagram	○	○	○	○
Neurological exam (e.g. dermatomes, myotomes, reflexes, etc.)	○	○	○	○
Neural dynamic testing (i.e. tests of neural mobility)	○	○	○	○
Neck Muscle Strength test (e.g. cervical isometric flexion, cervical isometric extension, etc.)	○	○	○	○
Neck muscle endurance testing (i.e. deep neck flexor endurance test)	○	○	○	○
Neck muscle stability testing (i.e. cranial-cervical flexion testing)	○	○	○	○
Posture alignment measures (i.e. dynamic analysis of scapular muscle control in posture and movement)	○	○	○	○
Upper extremity muscle strength/endurance (i.e. The Single Arm Military Press (SAMP) test)	○	○	○	○

Functional performance tests (i.e. Functional Impairment Test-Hand, and Neck/Shoulder/Arm (FIT-HaNSA)	○	○	○	○
Proprioception test (i.e. head and neck position sense (HNPS) testing)	○	○	○	○
Functional capacity assessment (e.g. timed weighted overhead test, timed supine capital flexion)	○	○	○	○

○ Other physical or functional measure/measures
Please specify:

Submit

Your responses have been submitted

Thank you for your time

If you have any queries regarding this survey or would like to hear about the results you can email Ahmad Alreni at a.alreni@shu.ac.uk .

Your responses are extremely valuable to us. We are looking for a UK sample of physiotherapists. if you know any other UK physiotherapists who are registered and practicing in the UK who could help with this survey, we would be grateful if you could forward on the link to this survey.

Yours sincerely,

Ahmad Alreni

Appendix 6

Survey Banner

UK National Survey of Neck Pain

If you are currently practicing in the UK and have seen at least one case of neck pain in the last 6 months, please complete this short survey.

Sheffield Hallam University | Centre for Health and Social Care Research

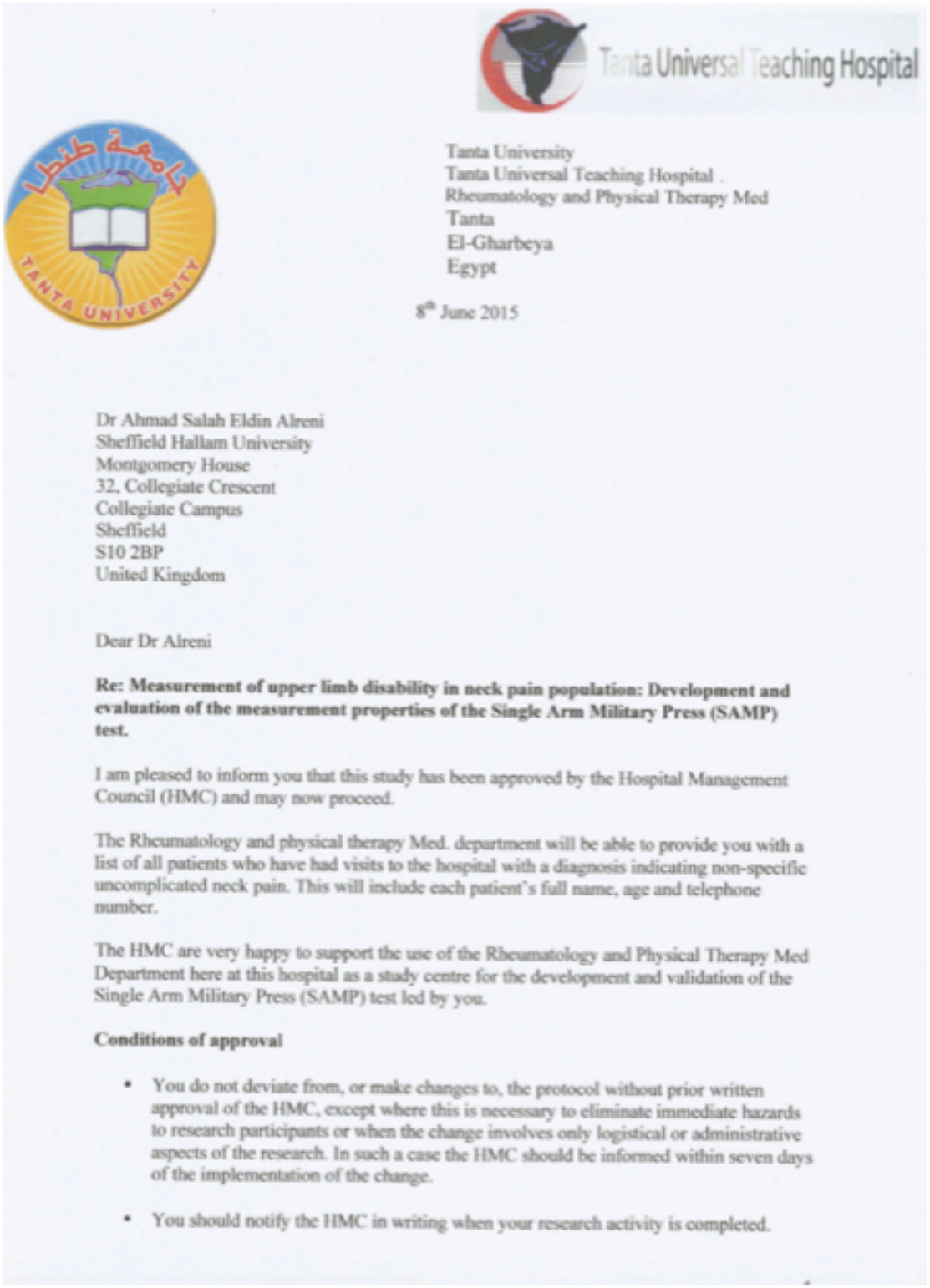
This survey is now OPEN, the aim of the survey is to investigate current physiotherapeutic management of neck pain throughout the UK.

The survey is aimed at UK PHYSIOTHERAPISTS who have been managing patients with neck pain. The findings will be used to understand current physiotherapy practice, to develop appropriate programme of research and to provide an overview of best practice.

Please help by completing the survey and share widely.

Appendix 7

Tanta Universal Teaching Hospital Ethical Approval (SAMP test Validation



Tanta Universal Teaching Hospital

Tanta University
Tanta Universal Teaching Hospital .
Rheumatology and Physical Therapy Med
Tanta
El-Gharbeya
Egypt

8th June 2015

Dr Ahmad Salah Eldin Alreni
Sheffield Hallam University
Montgomery House
32, Collegiate Crescent
Collegiate Campus
Sheffield
S10 2BP
United Kingdom

Dear Dr Alreni

Re: Measurement of upper limb disability in neck pain population: Development and evaluation of the measurement properties of the Single Arm Military Press (SAMP) test.

I am pleased to inform you that this study has been approved by the Hospital Management Council (HMC) and may now proceed.

The Rheumatology and physical therapy Med. department will be able to provide you with a list of all patients who have had visits to the hospital with a diagnosis indicating non-specific uncomplicated neck pain. This will include each patient's full name, age and telephone number.

The HMC are very happy to support the use of the Rheumatology and Physical Therapy Med Department here at this hospital as a study centre for the development and validation of the Single Arm Military Press (SAMP) test led by you.

Conditions of approval

- You do not deviate from, or make changes to, the protocol without prior written approval of the HMC, except where this is necessary to eliminate immediate hazards to research participants or when the change involves only logistical or administrative aspects of the research. In such a case the HMC should be informed within seven days of the implementation of the change.
- You should notify the HMC in writing when your research activity is completed.

- If you decide to terminate this research activity prematurely you send a report to the HMB within 15 days, indicating the reason for the early termination.

- You advise the HMC of any unusual or unexpected results that raise questions about the safety of the research.

- You provide the HMC with a summary of the results of this research and a copy of any future publication.

Approved documents

The final list of documents reviewed and approved by the HMC is as follows:

Application form
Research protocol
Questionnaires
Research participants information sheet
Research participants consent form
CV for the chief investigator "Dr Ahmad Alreni"
CV for the director of study "Professor Sionnadh McLean"

Yours sincerely

Prof Dr Ali Eldeeb

Dr. Ali Eldeeb
6/7/2015

Appendix 8

SHU Ethical Approval (SAMP test Validation)

Sheffield Hallam University | Centre for Health and Social Care Research

Date 26/10/2015

Research proposal number: 2014-5/ HWB-HSC-18

Dear Ahmad Alreni

This email relates to your research proposal:

Measurement of upper limb disability in neck pain population

I received your Major Amendment form on 26 October 2015. I note the change and believe it does not raise any substantial ethical issues requiring further review. I am therefore happy to approve the amendment. The relevant documents reviewed are amalgamated in a file:

Alreni amendment Binder.pdf

Good luck with your project.

Yours sincerely

Peter Allmark

Peter Allmark
Chair Faculty Research Ethics Committee
Faculty of Health and Wellbeing
Sheffield Hallam University
32 Collegiate Crescent
Sheffield
S10 2BP
0114 224 5727
p.allmark@shu.ac.uk

Appendix 9

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Telephone Checklist: Preliminary Screening (Phone Screening) English and Arabic

A list of patients who have had visits to the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital (Egypt) with a diagnosis indicating non-specific uncomplicated neck pain was obtained.

The chief investigator telephoned all prospective participants to answer the following questions as follows:

Hello, my name is Ahmad Alreni. I am phoning from Tanta Universal Teaching Hospital. We are working with Rheumatology and Physical Therapy Medicine Department to develop and evaluate health related outcome measure on patients suffering from neck pain.

Would you like to hear about it?

Recent research suggested that neck problems frequently associated with upper limb pain/disability. We want to develop a physical performance test (outcome measure) to evaluate the upper limb functional capacity and used in the assessment and during the management process of patients with neck pain. This will help us to identify and quantify any upper limb disability and cure it while treating neck pain, which will make neck pain patients to feel better and help them to cope better with normal daily activity.

So, we are recruiting people with neck pain and dividing them into three groups based on their age group, occupation, weight/height and the severity of their neck pain.

If you are interested in being involved in our study, you will be requested to attend one single assessment and testing session at Tanta Universal Teaching Hospital. In this session, you will be met by a member of our research team who will tell you more about the study, ask you to complete a questionnaire and carry out face-to-face assessment. If the testing procedure would be suitable and beneficial, you will be requested to perform a physical performance test for 30 seconds. This session will take up to 45-minute.

Would you be happy to participate?

- ☐ Yes
- ☐ No

If yes, may I ask you a few questions about your neck? This will help me to determine whether this testing procedure is suitable and good for you. Any information you give me will be kept confidential.

1. Have you had your neck symptoms for longer than 2 weeks? (*in weeks) (Acute/sub-acute or chronic pain)*
   - ☐ Yes
   - ☐ No

2. Have you had any treatment to your neck in the last 3 months?

   (What diagnosis? What the practitioner told you about your neck problem)

3. Are you planning to see anyone else for treatment?
   - Yes
   - No

If yes, to postpone until after the testing.

4. Do you have any other health problems such as dizziness, double vision, speech, swallowing, LOC (loss of consciousness)?
   - Yes
   - No

(Exclude patient with any major health problems)

5. Are you able to get on/off bed without help?
   - Yes
   - No

6. Are you able to walk, drive or use public transport without help?
   - Yes
   - No

7. Are you able to come to the new Tanta Universal Teaching Hospital?
   - Yes
   - No

Thank you for taking the time to answer those questions. I can tell you that at this stage it would seem that:

a. *The testing procedure would be suitable and beneficial for you (Make Appointment)*
b. *The testing procedure is not suitable for you (exclude this patient from the study).*

Making appointment (if a)
Can you come to Tanta Universal Teaching Hospital on (day) at (time) for face-to-face assessment, which be followed by the testing? *Negotiate an appointment time*.

Study information sheet
I would like to send you the study information sheet, where you will find more information about the study and the testing. Please read it and you may discuss your participation with your family and/or friends before attending the single face-to-face assessment and testing session: your address is…

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

استمارة تشخيص المريض

المرحلة الاولى من الفحص:

من خلال مكالمة هاتفية سوف يتم تعريف المريض بي وبمضمون البحث والغرض منه وما سبب اختياره ضمن عينة البحث لتوافر بعض الشروط متمثلة في الم الرقبة، وعند موافقته على الاشتراك في عينة البحث يطلب منه الاجابة على الاسئلة التالية:

١ – هل لديك أي الم في الرقبة لمدة اطول من اسبوعين؟

لا
نعم

٢ – هل عولجت خلال الثلاث شهور الماضية؟

لا
نعم (ما نوع التشخيص، ما هو تخصص المشخص)

٣ – هل تخطط لرؤية متخصص اخر بغرض العلاج قريبا؟

لا
نعم (يرجى التأجيل بعد اجراء الاختبار)

٤ – هل لديك أي مشاكل صحية اخرى مثل الدوار او ازدواج الرؤية او فقدان الوعي؟

لا
نعم (استبعاد أي مريض لدية مشاكل اخرى كبيرة)

٥- هل انت قادر على الحضور الى مستشفى طنطا التعليمي بدون مساعدة؟

لا (استبعاد أي مريض لا يستطيع الحضور لأسباب صحية)
نعم

نشكر الشخص المتحدث معنا لاستقطاعه جزء من وقته للإجابة على الاسئلة.
مع تحديد موعد لحضوره الى المستشفى التعليمي الجديد بطنطا في يوم ..... الساعة .....
معرفة العنوان الخاص به لإرسال معلومات أكثر عن البحث وإجراءاته واسباب اختياره في عينة البحث ومناقشة المشاركة في هذا البحث مع العائلة والاصدقاء المقربين قبل اتخاذ القرار النهائي في المشاركة.

Appendix 10

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Patient Information Sheet (English and Arabic)

We wish to invite you to participate in a research study. In order to have a clearer understanding of this research context, please read the following information sheet and do not hesitate to ask if there is anything that it not clear or you would like further information before you decide to take part.

What is the purpose of this study?

Neck pain is common, painful and many people report having trouble using their arm. Physiotherapists should try to measure the problems that people have when using their arm, so that they can advise patients how best to improve use of their arm.

The aim of this study is to develop and evaluate a physical performance test/performance-based outcome measure, the Single Arm Military Press (SAMP) Test, which is easy to use, economical, quick (maximum 2 minutes to perform and score). Preliminary research suggested that SAMP test is promising upper limb outcome measure for neck pain patients.

Why I have been invited?

We are inviting patients, age 18 years or over, with non-specific neck pain to take part in this study. The patients we are looking for should be able to travel to the Rheumatology and Physical Therapy department/clinic at Tanta Universal Teaching Hospital without support.

You have been invited to participate in this research program because of the type of neck pain you suffering from. If you are currently having treatment for your neck pain or had treatment in the past 3-month then you may be eligible for this study.

What will happen if I decide to take part?

If you are interested in being involved in this study, you will be asked to attend one single assessment and testing session, which will take up to 45 minutes. You will be requested to complete a questionnaire, which will give us information about your neck pain and your general physical and psychological well-being. This will be followed by face-to-face assessment to ensure that the testing procedure is suitable and beneficial for you. You will be then requested to sign a consent form to say that you agree to be involved with this study and complete the testing procedure.

It is preferable if you can wear a suitable, sleeveless/half-sleeves top, during the session so that the shoulder and elbow joints can be observed.

What do I have to do?

You will receive a complete demonstration and instruction of the test followed by warm-up and the test procedure, which will take up to 2-munites under direct supervision of a physiotherapist/physician.

Do I have to participate in this study?

'Only if you want to'

Participation is voluntary, you do not have to participate or you may withdraw from the study at any time before attending the face-to-face assessment and testing. However, please let us know if you are unable to participate at least 24-hours before your appointment. You do not need to tell us why you do not want to participate.

Are there any risks involved?
There are no known risks. This research program is simply validating performance-based outcome measure.

You may experience some muscle soreness, which is completely normal following physical exercise and may last up to 72 hours.

Confidentiality
All information from this study will be kept entirely confidential. All consent forms and any other identifiable information will be destroyed once the study has been completed. You will be informed of the results of the research if you wish.

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

معلومات شاملة لإجراءات البحث

نتشرف بدعوة سيادتكم بالمشاركة كجزء من عينة هذا البحث ولفهم ومعرفه المعلومات الشاملة لإجراءات البحث برجاء قراءة المعلومات التالية وعدم التردد في التساؤل عما إذا كان هناك أي معلومة غير واضحة او مزيد من المعلومات قبل ان تقرر المشاركة.

ما هو الغرض من هذه الدراسة؟

ان الم الرقبة من الامراض شائعة الحدوث والتي تؤدى الى الم شديد في بعض الاحيان، ومعظم مصابي الم الرقبة يجدون صعوبة في استخدام الذراع والكتف وبالأخص الاعمال التي تؤدى أعلى مستوى الرأس لذا يتوجب على الطبيب المتخصص في علاج الم الرقبة باختبار مدى الاعاقة في الكتف والذراع واليد.

الغرض من اجراء هذه الدراسة هو تطوير طريقة قياس الاعاقة في الكتف والذراع واليد عند مرضى الم الرقبة.

بعض الابحاث العلمية الحديثة اثبتت ان طريقة القياس (SAMP) هي طريقة مثالية للقياس.

لماذا تم اختياري في عينة البحث؟

لقد تم اختيارك كجزء من عينة البحث لأنك عمرك أكثر من ١٨ سنة وتعاني من الم في الرقبة.

ماذا سيحدث إذا قررت المشاركة في هذا البحث؟

إذا وافقت على المشاركة في هذه الدراسة فستدعى للحضور مرة واحدة في توقيت يتراوح بين ٣٠: ٤٥ دقيقة.

سوف يطلب منك الاجابة على بعض الاسئلة الخاصة بألم الرقبة وحالاتك الصحية العامة وملء بعض الاستبيانات.

ملحوظة: ان ترتدى المريضة تي شرت لرؤية الذراع كاملة.

ماذا على ان افعل في هذا الاختبار؟

سوف تتلقى تعليمات واضحة متبوعة بنموذج للاختبار يليه الاحماء ثم اداء الاختبار تحت اشراف متخصص.

هل من الضروري المشاركة في هذه الدراسة؟
(إذا اردت انت المشاركة)

المشاركة في هذه الدراسة تطوعية ويمكنك الانسحاب حينما تشاء.

هل هناك أي مخاطر يمكن ان تنتج عن المشاركة في هذه الدراسة؟

لا توجد مخاطر على الاطلاق ولكن قد تشعر ببعض الارهاق العضلي البسيط والذي هو مشابهة للذي يحدث بعد ممارسة الانشطة البدنية وقد يستغرق ذلك ٧٢ ساعة للاستشفاء.

الخصوصية:

جميع المعلومات الخاصة بالمرضى في هذه الدراسة هي سرية تامة وسوف يتم تدمير جميع اوراق الموافقة واي معلومات اخرى عن المرضى بمجرد الانتهاء من الدراسة.

Appendix 11

Neck Disability Index (NDI) English and Arabic

This questionnaire has been designed to give us information as to how your neck pain has affected your ability to manage in everyday life. Please answer every section and mark in each
section only the one box that applies to you. We realise you may consider that two or more statements in any one section relate to you, but please just mark the box that most closely describes your problem.

Section 1: Pain Intensity

- I have no pain at the moment
- The pain is very mild at the moment
- The pain is moderate at the moment
- The pain is fairly severe at the moment
- The pain is very severe at the moment
- The pain is the worst imaginable at the moment

Section 2: Personal Care (Washing, Dressing, etc.)

- I can look after myself normally without causing extra pain
- I can look after myself normally but it causes extra pain
- It is painful to look after myself and I am slow and careful
- I need some help but can manage most of my personal care
- I need help every day in most aspects of self-care
- I do not get dressed, I wash with difficulty and stay in bed

Section 3: Lifting

- I can lift heavy weights without extra pain
- I can lift heavy weights but it gives extra pain
- Pain prevents me lifting heavy weights off the floor, but I can manage if they are conveniently placed, for example on a table
- Pain prevents me from lifting heavy weights but I can manage light to medium weights if they are conveniently positioned
- I can only lift very light weights
- I cannot lift or carry anything

Section 4: Reading

- I can read as much as I want to with no pain in my neck
- I can read as much as I want to with slight pain in my neck
- I can read as much as I want with moderate pain in my neck
- I can’t read as much as I want because of moderate pain in my neck
- I can hardly read at all because of severe pain in my neck
- I cannot read at all

Section 5: Headaches

- I have no headaches at all
- I have slight headaches, which come infrequently
- I have moderate headaches, which come infrequently
- I have moderate headaches, which come frequently
- I have severe headaches, which

Section 6: Concentration

- I can concentrate fully when I want to with no difficulty
- I can concentrate fully when I want to with slight difficulty
- I have a fair degree of difficulty in concentrating when I want to
- I have a lot of difficulty in concentrating when I want to

come frequently
- I have headaches almost all the time

- I have a great deal of difficulty in concentrating when I want to
- I cannot concentrate at all

Section 7: Work

- I can do as much work as I want to
- I can only do my usual work, but no more
- I can do most of my usual work, but no more
- I cannot do my usual work
- I can hardly do any work at all
- I can't do any work at all

Section 8: Driving

- I can drive my car without any neck pain
- I can drive my car as long as I want with slight pain in my neck
- I can drive my car as long as I want with moderate pain in my neck
- I can't drive my car as long as I want because of moderate pain in my neck
- I can hardly drive at all because of severe pain in my neck
- I can't drive my car at all

Section 9: Sleeping

- I have no trouble sleeping
- My sleep is slightly disturbed (less than 1 hr sleepless)
- My sleep is mildly disturbed (1-2 hrs sleepless)
- My sleep is moderately disturbed (2-3 hrs sleepless)
- My sleep is greatly disturbed (3-5 hrs sleepless)
- My sleep is completely disturbed (5-7 hrs sleepless)

Section 10: Recreation

- I am able to engage in all my recreation activities with no neck pain at all
- I am able to engage in all my recreation activities, with some pain in my neck
- I am able to engage in most, but not all of my usual recreation activities because of pain in my neck
- I am able to engage in a few of my usual recreation activities because of pain in my neck
- I can hardly do any recreation activities because of pain in my neck
- I can't do any recreation activities at all

مقياس إعاقة الرقبة (NDI)

تستفسر هذه السلسلة من الاسئلة عن الحالات او الاعراض التي تشعر بها في رقبتك وعن قدرتك على تأدية نشاطات معينة. الرجاء ان تجيب على كل سؤال بناء على حالتك بوضع علامة في المربع المقابل الى الاجابة الاقرب الى الدقة

المجموعة الاولى (شدة الام الرقبة)

لا يوجد الم في الوقت الحالي	
يوجد الم بشكل بسيط جدا	
يوجد الم بشكل بسيط	
يوجد الم شديد	
يوجد الم شديد جدا	
يوجد الم شديد بدرجة كبيرة	

المجموعة الثانية (العناية الشخصية مثل خلع الملابس، الاستحمام ..... الخ)

يمكنني ان اعتنى بأموري الشخصية بدون زيادة في الالم	
يمكنني ان اعتنى بأموري الشخصية ولكنة يسبب الام اضافية	
من المؤلم الاعتناء بأموري الشخصية وعند قيامي اكون بطيء وحريص	
احتاج الى بعض المساعدة في بعض الامور واغلب الامور الاخرى اقوم بها	
انا احتاج الى المساعدة في معظم امور العناية الشخصية	
لا أستطيع العناية بأموري الشخصية مثل ارتداء الملابس واغسل وجهي بصعوبة وأبقى في السرير طول الوقت	

المجموعة الثالثة (رفع الاشياء باليد)

أستطيع رفع الاوزان الثقيلة بدون الم إضافي	
أستطيع رفع الاوزان الثقيلة ولكن يسبب الم إضافي	
الالم الرقبة تمنعني من رفع الاوزان الثقيلة من الارض ولكن حملها إذا كانت على ارتفاع مناسب (موجودة على طرابيزة)	
ا الام الرقبة تمنعني من رفع الاوزان الثقيلة من الارض ولكن أستطيع رفع الاوزان الخفيفة والمتوسطة إذا كانت على ارتفاع مناسب (موجودة على طرابيزة)	
أستطيع رفع الاوزان الخفيفة جدا فقط	
لا أستطيع رفع او حمل أي شيء	

المجموعة الرابعة (القراءة)

أستطيع القراءة بقدر ما اريد بدون الام في الرقبة	
أستطيع القراءة بقدر ما اريد مع وجود الام خفيفة في الرقبة	
أستطيع القراءة بقدر ما اريد مع وجود الام متوسطة في الرقبة	
لا أستطيع القراءة بقدر ما اريد بسبب وجود الام متوسطة في الرقبة	
أستطيع القراءة بصعوبة بسبب وجود الام عالية في الرقبة	
لا أستطيع القراءة على الاطلاق	

المجموعة الخامسة (الصداع)

لا يوجد لدى صداع على الاطلاق	
لدى صداع خفيف ويأتي بشكل غير منتظم	
لدى صداع متوسط ويأتي بشكل غير منتظم	
لدى صداع متوسط ويأتي بشكل منتظم	
لدى صداع عالي ويأتي بشكل منتظم	
لدى صداع في كل الاوقات تقريبا	

المجموعة السادسة (التركيز)

يمكنني التركيز بشكل عالي عندما اريد مع عدم وجود أي صعوبة	
يمكنني التركيز بشكل عالي عندما اريد مع وجود صعوبة بسيطة	
لدى صعوبة متوسطة في التركيز	
لدى صعوبة عالية في التركيز	
لدى صعوبة كبيرة جدا في التركيز	
لا أستطيع التركيز على الاطلاق	

المجموعة السابعة (العمل)

يمكنني القيام بالعمل وبأي اعمال اضافية اخرى	
يمكنننني القيام بالعمل المعتاد فقط	
أستطيع القيام بمعظم العمل المعتاد فقط	
لا يمكنننني القيام بالعمل المعتاد	
أستطيع بصعوبة جدا القيام بأي عمل	
لا أستطيع القيام بأي عمل على الاطلاق	

المجموعة الثامنة (القيادة)

أستطيع قيادة السيارة بدون أي الالم في الرقبة	
أستطيع قيادة السيارة مع وجود الالم خفيفة في الرقبة	
أستطيع قيادة السيارة مع وجود الالم متوسطة في الرقبة	
لا أستطيع قيادة السيارة بسبب وجود الالم متوسطة في الرقبة	
أستطيع بصعوبة قيادة السيارة بسبب وجود الالم عالية في الرقبة	
لا أستطيع قيادة السيارة على الاطلاق	

المجموعة التاسعة (النوم)

ليس لدى أي مشكلة في النوم	
اشعر بالقلق البسيط لمدة اقل من ساعة اثناء النوم	
اشعر بالقلق المتوسط الذي يتراوح بين ١ ساعة: ٢ ساعة اثناء النوم	
اشعر بالقلق المتوسط الذي يتراوح بين ٢ ساعة: ٣ ساعة اثناء النوم	
اشعر بالقلق العالي الذي يتراوح بين ٣ ساعة: ٥ ساعة اثناء النوم	
اشعر بالقلق الشديد الذي يتراوح بين٥ ساعة :٧ ساعة اثناء النوم	

المجموعة العاشرة (الانشطة الاجتماعية والترفيهية)

أستطيع المشاركة في جميع الانشطة الاجتماعية والترفيهية بدون الالم في الرقبة على الاطلاق	
أستطيع المشاركة في جميع الانشطة الاجتماعية والترفيهية مع بعض الالم في الرقبة	
أستطيع المشاركة في معظم وليس جميع الانشطة الاجتماعية والترفيهية بسبب الالم في الرقبة	
أستطيع المشاركة في قليل من الانشطة الاجتماعية والترفيهية بسبب الالم في الرقبة	
أستطيع بصعوبة المشاركة في قليل من الانشطة الاجتماعية والترفيهية بسبب الالم في الرقبة	
لا أستطيع المشاركة في الانشطة الاجتماعية والترفيهية على الاطلاق	

Appendix 12

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Face to Face Assessment (English and Arabic)

Participant's demographic information

Assessment date	--/--/----
Participant name	
Participant ID number (in the study)	
Date of Birth	--/--/----
Occupation	
Weight	
Height	
Telephone number	
Email address	
Home address	

- Where is your symptoms now? *Complete the body chart*.

- How severe is your symptoms now?

  a. Neck pain/symptoms
  No pain 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 worst possible pain (*see Appendix 11*)

  b. Upper limb pain/symptoms
  No pain 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 worst possible pain (*see Appendix 11*)

- Over the past 2 weeks are your symptoms:
  - Getting better
  - Getting worse
  - Same

If worsening, in what way? (exclude deteriorating neurological condition e.g. cord sign, radiculopathy)

- Were you involved in an accident which caused your pain? (exclude recent major trauma)

- How long have you had your neck symptoms? In weeks 'determine whether acute, sub-acute or chronic'

- How long ago did you first experience these symptoms? In weeks 'determine recurrence' and how it was treated?

- Have you had any treatment such as physical therapy in the past three months?

- Have you ever injured your shoulder, arm or hand substantially? exclude injuries which has resulted in current or prolonged disability
- Is your weight steady? Exclude unintentional weight loss

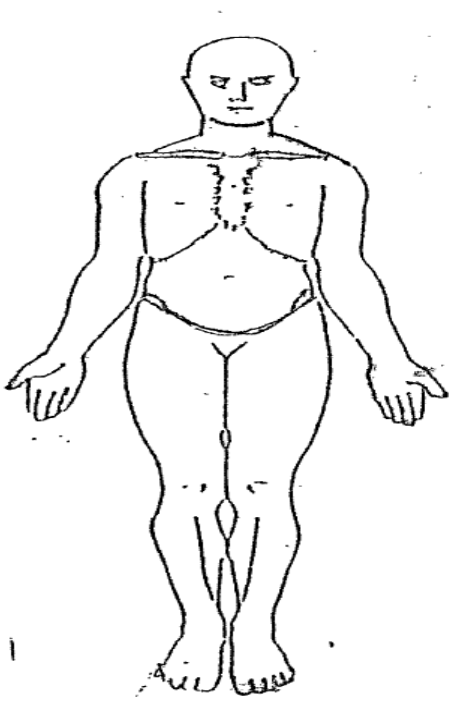
- Are you sleeping OK at night? Exclude severe pain at night

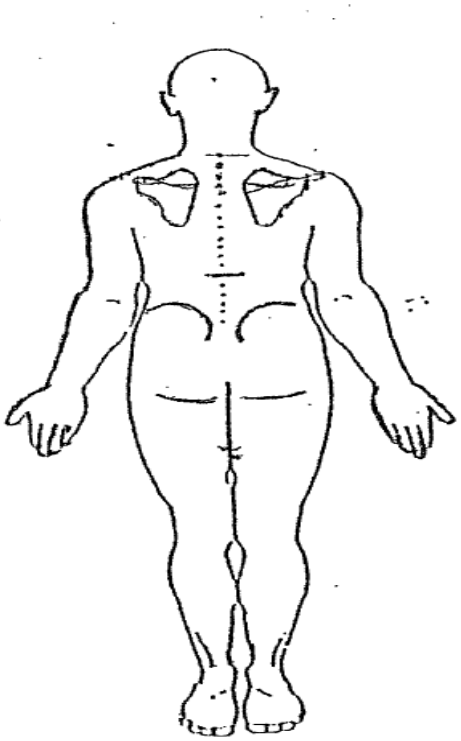
- Are you having any problems with dizziness, double vision, speech, swallowing, LOC (loss of consciousness)? Exclude in accordance with the criteria

- Do you have any general medical problems? If so, specify what type of problems? (Exclude vertebral artery problems)

- Do you have any general medical problems? If so specify, (Exclude – severe rheumatoid arthritis, severe multiple sclerosis, cancer, osteoporosis, cardiac conditions, severe SOBOE, uncontrolled hypertension, postural hypotension, balance problems).

If all answers confirmed eligibility and the NDI scored at least 10%, and patient still happy to proceed. Patient should sign the consent form and allocated for SAMP testing.





Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

المرحلة الثانية من الفحص (بحضور المريض):

تاريخ يوم الفحص	...../...../.....
اسم المريض	
رقم المريض	
تاريخ ميلاد المريض	
وظيفة المريض	
الوزن	
الطول	
رقم تليفون المريض	
عنوان المريض	
مرحلة البحث	

السؤال الاول:
اين يوجد الالم الان؟ (استكمال مخطط الجسم المرفق)

السؤال الثاني:
ما مدى قوة الالم؟

الم الرقبة:
لا يوجد الم 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 اقصى الم ممكن

الم الذراع والكتف واليد:
لا يوجد الم 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 اقصى الم ممكن

السؤال الثالث:
خلال الاسبوعين الماضيين الم الرقبة:
تحسن:
ازداد سوء: (استبعاد الحالات المرتبطة بالعمود الفقري والاعصاب)
نفس الألم:

السؤال الرابع:
هل الم الرقبة الذي تشعر به نتيجة لحادث؟
لا:

نعم: (استبعاد المرضى المصابون حديثا)

السؤال الخامس:
ما مدة شعورك بألم الرقبة في الوقت الراهن؟ (بالأسبوع)

السؤال السادس:
متى شعرت بألم الرقبة للمرة الاولى؟ (بالأسبوع)
.....

السؤال السابع:
هل تلقيت أي علاج في الثلاث شهور الماضية؟
لا:
نعم:

السؤال الثامن:
هل سبق لك ان تعرضت لإصابة في الكتف او الذراع او اليد؟
لا:
نعم: (استبعاد المصابون بالألم شديد او إعاقة في الكتف او الذراع او اليد)

السؤال التاسع:
هل وزنك ثابت؟
لا: (استبعاد أي نقصان في الوزن بدون قصد)
نعم:

السؤال العاشر:
هل تنام ليلا بدون اضرابات؟
نعم:
لا: (استبعاد الالم الليلي)

السؤال الحادي عشر:
هل لديك أي مشاكل صحية مثل الدوار، ازدواج الرؤية، او فقدان الوعي؟
لا
نعم: (استبعاد المرضى مع هذه المشكلات الطبية)

إذا كانت الاجابات مطابقة لأسس اختيار المريض ونتيجة مقياس إعاقة الرقبة (NDI) ١٠ ٪ على الاقل يتم توقيع المريض على موافقته للاشتراك في عينة البحث واجراءه.

Appendix 13

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Consent Form (English and Arabic)

Name of the researcher: Ahmad Alreni

1. I confirm that I have read and understand the information sheet dated 0 /0 /00 for the above research study and have had opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason. ☐
3. I agree to tack part in the above research study ☐

Name of the subject (BLOCK CAPITALS):

……

Signature:
……Date:
……

Signature of the researcher:

……

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

موافقة المريض على الاشتراك في عينه البحث

اسم الباحث: احمد صلاح الدين العريني

١. اقر انا الموقع ادناه على إنني قرأت وفهمت المعلومات الخاصة بإجراءات هذا البحث وكان لدى المقدرة على الاستفسار عن بعض الاجزاء الغامضة لدى في اجراءات البحث وتم الاجابة على استفساراتي

٢. اتفهم انا الموقع ادناه إنني اشارك تطوعيا في هذا البحث مع مقدرتي في الانسحاب منه بدون ابداء أي اسباب

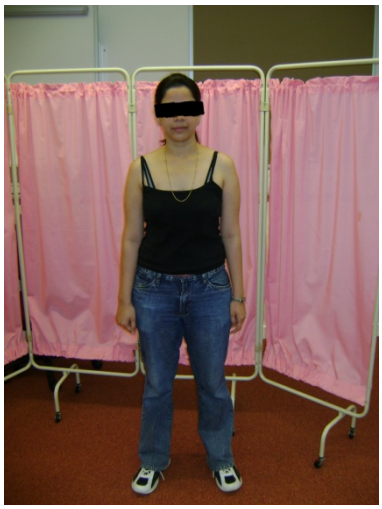
٣. اوافق على الاشتراك في عينة هذا البحث

اسم المريض: .....
التوقيع: .....
التاريخ: .....

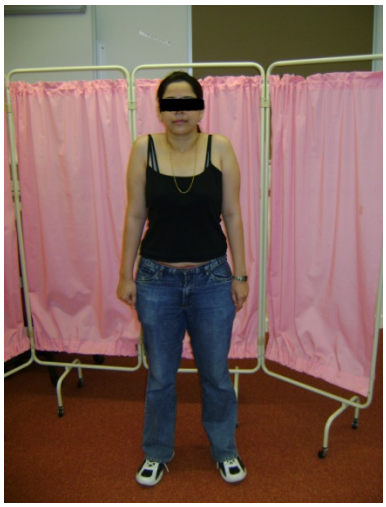
Appendix 14

Brief Warm-Up before SAMP Testing

Shoulder Shrugs (10 reps)

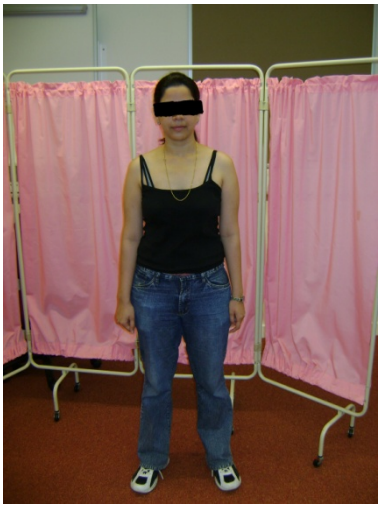


Starting position

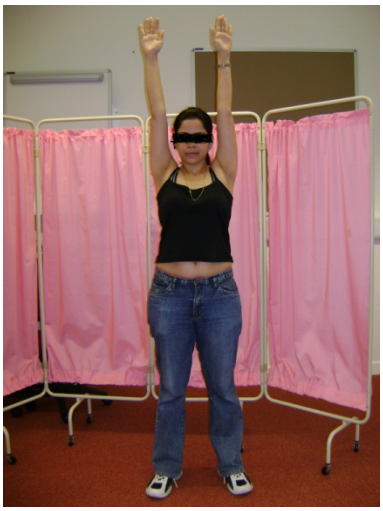


End position

Shoulder Flexion (10 reps)

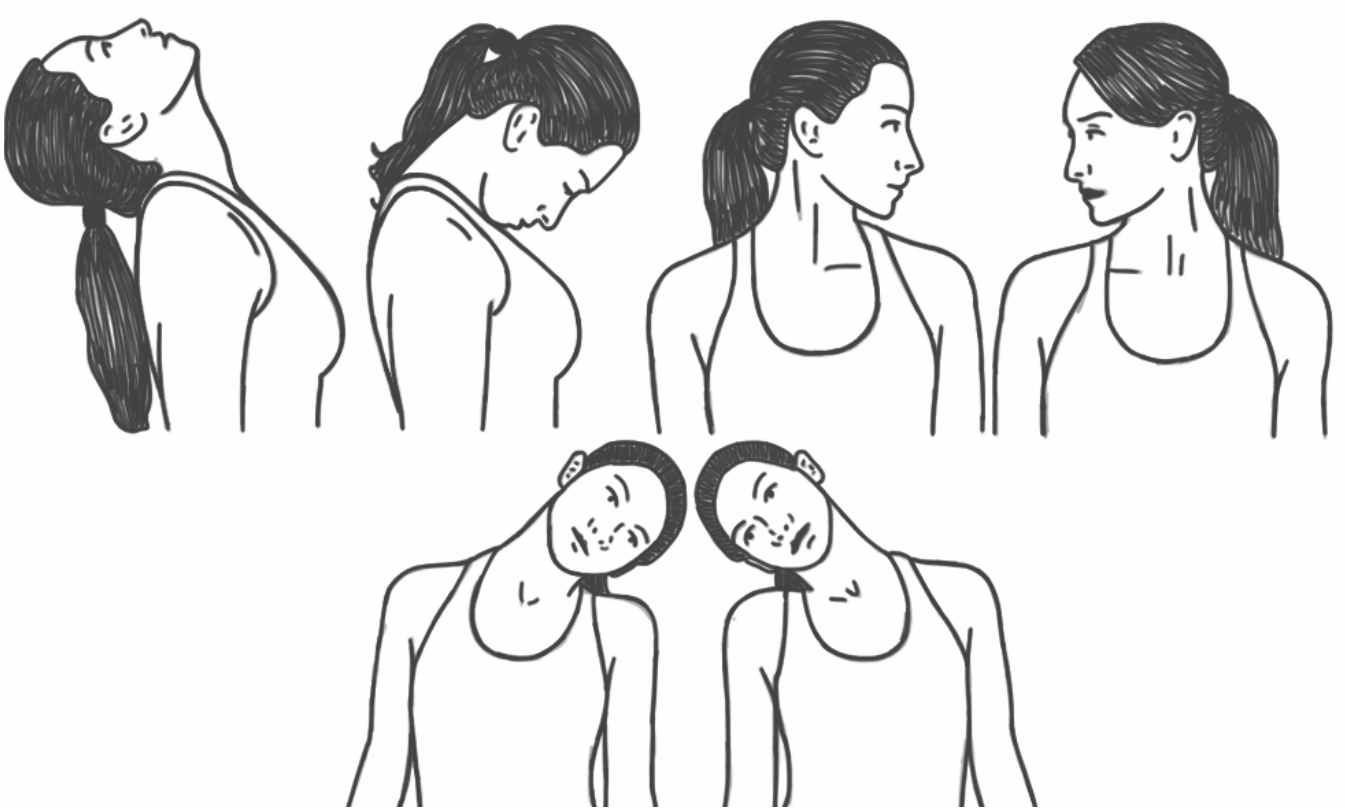


Starting position

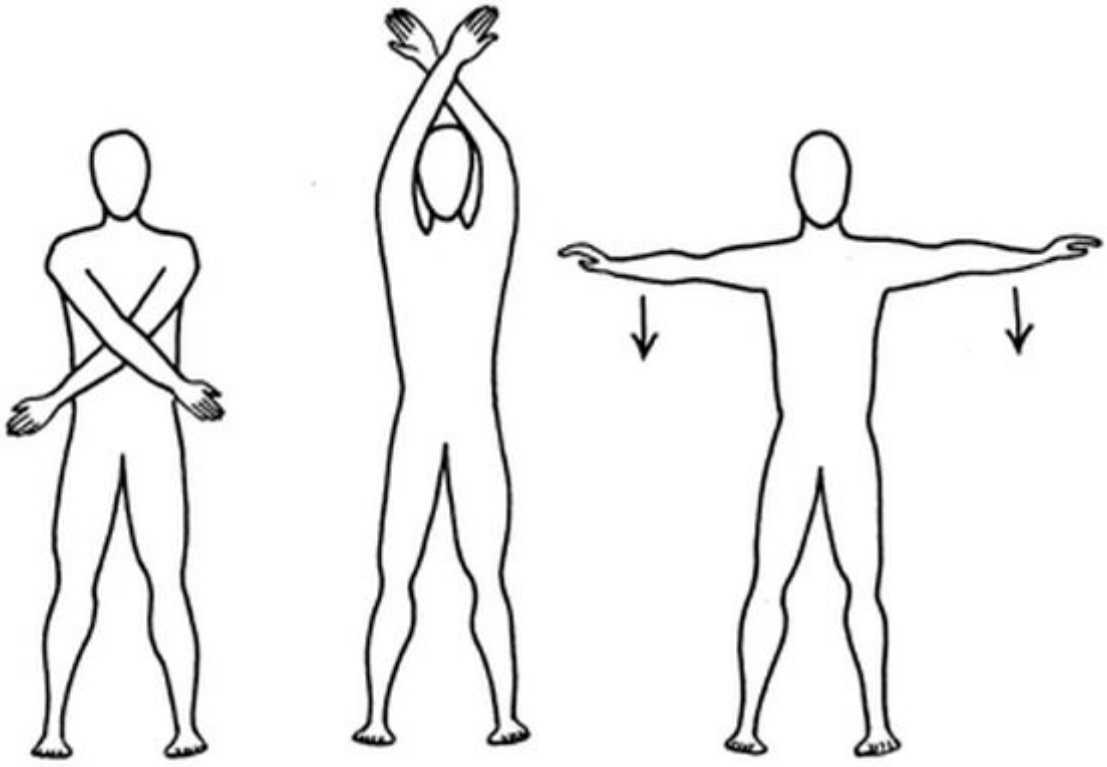


End position

Range of Motion Exercise for the Neck



Range of Motion Exercise for the Shoulder



Appendix 15

SAMP Test Procedure

Description and practical application of the SAMP test

The SAMP test is a performance-based instrument that would be used for diagnostic purpose to measure the physical functioning, capacity, of the upper limb on a specific targeted population (patients with non-specific neck pain). The International Classification of Functioning, Disability and Health Categories Code is d430-d449: carrying, moving and handling object.

The SAMP test is designed to assess the strength and endurance of the upper limb by counting the number of repetitions a patient can perform in 30 seconds. This is to enable clinicians in the field of neck pain to assess wide variations pertaining to the patients' ability level with the possible scoring range between zero for those who cannot complete even one repetition, which indicate high level of pain/disability, to a high of 30 or more for highly fit individual (McLean et al., 2010a). In addition, the administrative and respondent burden of the SAMP test protocol are minimal < 1 minute, no formal instruction to obtain and the equipment required are only dumbbell and timer/stop watch.

Test Equipment

½-kg, 1-kg and 1½-kg hand weight

Stop watch

SAMP test procedure

The SAMP test is conducted in the standing position with the feet at the shoulder width, patient is requested to carry a dumbbell and lifted using their dominant hand to the shoulder level (see Starting Position). Patient is requested to repeat raising the hand with the dumbbell directly overhead by extending through the elbow (see End Position) and repeat this process as fast as possible for 30 seconds.

Starting position

End position

SAMP test standardised verbal instructions

"For the purpose of this test, please do the best you can by raising your hand with the dumbbell overhead as fast as you can but do not push yourself beyond what you think is safe for you.

1. Stand erect with your feet flat in the floor and at the shoulder width apart with the dumbbell at the shoulder level.
2. On the signal to begin, raise your hand with the dumbbell overhead a full and then come back to the shoulder level.
3. Keep going for 30 seconds and until I say stop
4. Get ready and start"

Test Stopping Criteria

The SAMP test should be continued for 30 seconds, but is terminated based on the following stopping rules:

1. Participant stops or states it is too painful to continue.
2. Participant is severely off pacing to the extent that they are unable to complete one repetition of the movement.
3. Participant substitutes using trunk/whole body movement and cannot correct with feedback.
4. The examiner believes that the participant is at risk of injury or adverse complication if test is to continue.

Scoring:

The number of valid (correct) repetitions with the 30 seconds.

Appendix 16

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Data Collection Sheet (English and Arabic)

Assessment Date:
Examiner Name:
Group Number:
Weight used:

Patient ID Number	SAMP Score reps/30-sec	Administration Time	Completion Time

SAMP score = number of repetitions in the 30-second, **Administration time** = description, demonstration, instructions and the 30-sec performance, **Completion time** = warm-up time and administration time.

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

استمارة جمع البيانات

تاريخ اداء الاختبار: .....
اسم الفاحص: .....
رقم المجموعة: .....
الوزن المستخدم: .....

رقم المريض	ت \ ٣٠ث	زمن الاختبار	الزمن الكلى الاختبار

- ت\ ٣٠ ث = عدد التكرارات على ٣٠ ثانية
- زمن الاختبار = شرح الاختبار، اداء نموذج، التعليمات + ٣٠ ث اداء الاختبار
- الزمن الكلى للاختبار = الاحماء + زمن الاختبار

Appendix 17

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Symptoms Severity Scale

Severity	Description of Experience
10 - Worst Possible Pain	I am in bed and can't move due to my pain. I need someone to take me to emergency room to get help for my pain.
9 – Extremely Sever Pain	My Pain is all that I can think about. I can barely talk or move because of the pain.
8 – Very Severe Pain	My pain is so severe that it is hard to think of anything else. Talking and listening are difficult.
7 – Severe Pain	I am in pain all the time. It keeps me from doing most activities.
6 – Distressing Pain	I think about my pain all the time. I give up many activities because of my pain.
5 – Distracting Pain	I think about my pain most of the time. I cannot do some of the activities I need to do each day because of the pain.
4 – Moderate pain	I am constantly aware of my pain but I can continue most activities.
3 – Uncomfortable Pain	My pain bothers me but I can ignore it most of the time.
2 – Mild Pain	I have a low level of pain. I am aware of my pain only when I pay attention to it.
1 – Slight Pain	My pain is hardly noticeable.
0 – No Pain	I have no pain

https://paindoctor.com/pain-scales/

Appendix 18

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Patient Input after Testing (English and Arabic)
(To be completed for each patient immediately after testing by the researcher)

1. Now I am going to ask you about your symptoms after performing the SAMP test:
    - How severe is your neck pain/symptoms now?

0	1	2	3	4	5	6	7	8	9	10

0=No pain, 1=Slight pain, 2=Mild pain, 3=Uncomfortable pain, 4=Moderate pain, 5=Distracting pain, 6=Distressing pain, 7=Severe pain, 8=Very severe pain, 9=Extremely severe pain, 10=The Worst Possible pain/Symptoms

    - How severe is your upper limb pain/symptoms now?

0	1	2	3	4	5	6	7	8	9	10

0=No pain, 1=Slight pain, 2=Mild pain, 3=Uncomfortable pain, 4=Moderate pain, 5=Distracting pain, 6=Distressing pain, 7=Severe pain, 8=Very severe pain, 9=Extremely severe pain, 10=The Worst Possible pain/Symptoms

2. Now I am going to ask you about your experience with the SAMP testing:
    - How light or heavy was the dumbbell you used in the SAMP testing procedure?

1	2	3	4	5	6	7	8	9

1=Extremely Light, 2=Very Light, 3=Moderately Light, 4= Slightly Light, 5=Neither light nor Heavy, 6=Slightly Heavy, 7=Moderately Heavy, 8=Very Heavy, 9=Extremely Heavy

    - How easy or difficult was it to understand the instruction and perform the SAMP test?

1	2	3	4	5	6	7	8	9

1=Extremely Easy, 2=Very Easy, 3=Moderately Easy, 4=Slightly Easy, 5=Neither Difficult nor Easy, 6=Slightly Difficult, 7=Moderately Difficult, 8=Very Difficult, 9=Extremely Difficult.

    - How easy or difficult was the SAMP test in relation to your ability to perform the procedure?

1	2	3	4	5	6	7	8	9

1=Extremely Easy, 2=Very Easy, 3=Moderately Easy, 4=Slightly Easy, 5=Neither Difficult nor Easy, 6=Slightly Difficult, 7=Moderately Difficult, 8=Very Difficult, 9=Extremely Difficult.

    - How suitable or unsuitable was the SAMP test with regard to the time and effort used?

1	2	3	4	5	6	7	8	9

1=Highly Suitable, 2=Considerably Suitable, 3=Fairly Suitable, 4=Slightly Suitable, 5=Neither Suitable nor Unsuitable, 6=Slightly Unsuitable, 7=Fairly Unsuitable, 8=Considerably Unsuitable, 9=Completely Unsuitable.

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

رأى المريض

- ما هي درجه الام الرقبة الان؟

٠	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠

٠=لا يوجد الم، ١٠= اعلي درجه من الألم

ما هي درجه الام الكتف واذراعين الان؟

٠	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠

٠=لا يوجد الم، ١٠= اعلي درجه من الألم

- ماهي درجة مناسبة/سهوله الثقل المستخدم؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

- ماهي درجة مناسبة/صعوبة الاختبار بالنسبة للتعليمات والأداء؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

- ماهي درجة مناسبة الاختبار من جانب مقدرتك على ادا الاختبار؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

- ماهي درجة مناسبة/صعوبة الاختبار بالنسبة للوقت المستخدم والمجهود المبذول؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

Appendix 19

Measuring upper limb disability in neck pain population: evaluation of the acceptability and feasibility of the Single Arm Military Press (SAMP) test

Examiner Input after testing (English and Arabic)
(To be completed once only after the completion of the SAMP testing)

Do you feel that:

	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
There is a need for extra training to understand the application of the SAMP test procedure					
There is a need for extra staff to support the application of the SAMP test procedure					
There is a need for any technological support when using the SAMP test for patients with neck pain					

1. How easy or difficult was it to provide explanation with demonstration of the SAMP test to patients?

1	2	3	4	5	6	7	8	9

1=Extremely Easy, 2=Very Easy, 3=Moderately Easy, 4=Slightly Easy, 5=Neither Difficult nor Easy, 6=Slightly Difficult, 7=Moderately Difficult, 8=Very Difficult, 9=Extremely Difficult.

2. How easy or difficult was it regarding the overall administration of the SAMP test?

1	2	3	4	5	6	7	8	9

1=Extremely Easy, 2=Very Easy, 3=Moderately Easy, 4=Slightly Easy, 5=Neither Difficult nor Easy, 6=Slightly Difficult, 7=Moderately Difficult, 8=Very Difficult, 9=Extremely Difficult.

3. How appropriate or inappropriate was the SAMP test regarding resources needed (e.g. time, cost)?

1	2	3	4	5	6	7	8	9

1=Highly appropriate, 2=Considerably Appropriate, 3=Fairly Appropriate, 4=Slightly Appropriate, 5=Neither appropriate nor Inappropriate, 6=Slightly Inappropriate, 7=Fairly Inappropriate, 8=Considerably Inappropriate, 9=Completely Inappropriate.

رأى الفاحص

	أوافق بقوه	اوافق	لا ارفض ولا اوافق	ارفض	ارفض بقوه
هناك احتياج الي تدريب خاص من اجل استخدام هذا الاختبار					
هناك احتياج مساعده من شخص اخر لاستخدام هذا الاختبار					
هناك احتياج الي مساعده تكنولوجيه من اجل استخدام هذا الاختبار					

ماهي درجة مناسبة الاختبار/صعوبة الاختبار من جانب التعليمات واداء النموذج؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

ماهي درجة مناسبة/صعوبة الاختبار من جانب التنفيذ وحساب النتيجة؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

ماهي درجة مناسبة/صعوبة الاختبار من جانب الوقت المستخدم والتكلفة المالية؟

١	٢	٣	٤	٥	٦	٧	٨	٩

١=اعلي درجه من السهولة، ٩=اعلي درجه من الصعوبة

Appendix 20

Descriptive Data for all Measures for SAMP Acceptability and Feasibility

Group 1 SAMP Testing Using (½kg)

Patient ID	SAMP Admin Time / Completion Time /Seconds	SAMP Scores Mean= 20.7	NDI Score	Neck Pain/ symptoms			Upper limb Pain/symptoms			Patient Input (weight)	Patient Input (Instruction/ Performance	Patient Input (Ability to Perform)	Patient Input (Time & Effort)
				Before Testing	Immediately After testing	24 H After Testing	Before Testing	Immediately After Testing	24 H After Testing				
1	60/120	32	16%	3	3	2	1	3	1	1	1	1	1
5	60/120	26	30%	4	5	3	1	4	1	1	1	1	1
7	60/120	30	25%	4	4	2	1	3	1	1	1	1	1
11	57/117	22	35%	4	4	3	1	4	1	1	1	1	1
13	58/118	26	32%	4	5	4	1	3	1	1	1	1	1
16	60/120	28	41%	5	5	3	1	4	1	1	1	1	1
19	55/115	17	50%	5	8	5	3	5	3	1	2	1	1
24	60/120	30	24%	4	5	3	2	5	1	1	1	1	1
27	60/120	26	23%	4	6	4	2	4	1	1	1	1	1
30	60/120	33↑	15%	3	5	3	1	4	1	1	1	1	1
31	58/118	20	17%	3	4	3	2	3	2	1	2	1	1
32	57/117	20	19%	3	4	3	1	3	1	1	1	1	1
33	56/116	19	30%	4	5	3	2	3	1	1	1	1	2
34	58/118	21	30%	4	5	3	2	3	2	1	1	1	1
35	54/114	16	35%	4	5	3	2	4	2	2	1	4	2
36	50/110	11	52%	5	6	4	3	5	3	3	1	5	3
37	50/110	15	36%	4	5	3	2	4	2	2	1	5	3
38	50/110	11	55%	5	6	4	3	5	3	4	1	5	3
39	50/110	14	34%	4	5	3	3	4	3	2	1	4	3
40	55/115	16	30%	4	5	3	3	5	3	2	1	4	3
41	50/110	10↓	75%	6	7	6	4	6	4	3	1	5	4
42	52/112	15	50%	4	5	3	3	4	2	2	1	3	3
43	50/110	15	58%	4	5	4	3	4	3	2	1	3	3

SAMP: Single Arm Military Press, NDI: Neck and Disability Index, Admin Time: Administration Time (Description, demonstration, instruction and 30S Performance). Completion Time (Warm-Up Time & Administration Time). ↓=Minimum Score. ↑=Maximum Score.

Group 2 SAMP Testing Using (1-kg)

Patient ID	SAMP Admin Time / Completion Time /Seconds	SAMP Scores Mean =15.88	NDI Score	Neck Pain/ symptoms			Upper limb Pain/symptoms			Patient Input (weight)	Patient Input (Instruction / Performance)	Patient Input (Ability to Perform)	Patient Input (Time & Effort)
				Before Testing	Immediately After Testing	24 Hours After Testing	Before Testing	Immediately After Testing	24 Hours After Testing				
3	60/60	24	27%	4	5	3	1	4	1	3	1	3	2
6	60/60	23	36%	4	5	3	1	3	1	3	1	3	2
8	60/60	22	29%	5	6	3	1	4	1	3	1	2	2
10	55/60	17	48%	5	6	4	3	6	3	4	1	3	3
14	60/60	25	28%	4	5	3	1	5	1	3	1	2	2
17	60/60	26	25%	4	5	3	1	4	1	3	1	2	2
21	60/60	26	28%	5	7	4	2	5	2	3	1	2	2
22	60/60	24	35%	5	6	4	2	5	1	3	1	2	3
25	60/60	25	26%	4	6	3	2	5	1	3	1	2	2
29	60/60	27↑	18%	3	4	2	1	4	1	2	1	2	1
44	50/60	7	80%	6	7	5	4	7	4	5	1	6	6
45	50/60	7	70%	5	6	4	3	6	3	6	1	7	6
46	50/60	11	48%	4	6	3	3	5	3	4	1	4	3
47	50/60	13	45%	3	5	3	2	5	2	4	1	4	3
48	53/60	12	40%	4	5	3	2	4	2	4	1	5	2
49	50/60	9	60%	5	6	4	3	5	3	5	1	6	4
50	54/60	11	45%	4	5	3	3	4	3	5	1	6	4
51	50/60	6↓	65%	5	7	5	4	7	4	7	1	7	6
52	55/60	14	39%	4	6	4	2	4	2	4	1	4	2
53	55/60	11	58%	5	6	4	3	5	3	4	1	5	3
54	55/60	11	60%	5	6	4	3	5	3	5	1	5	4
55	53/60	14	40%	4	5	3	2	4	2	4	1	5	2
56	50/60	9	65%	5	7	5	4	7	4	5	1	4	3
57	50/60	7	68%	5	6	4	4	7	4	6	1	6	5

SAMP: Single Arm Military Press, NDI: Neck and Disability Index, Admin Time: Administration Time (Description, demonstration, instruction and 30S Performance). Completion Time (Warm-Up Time & Administration Time). ↓=Minimum Score. ↑=Maximum Score.

Group 3 SAMP Testing Using (1½-kg)

Patient ID	SAMP Admin Time / Completion Time /Seconds	SAMP Scores Mean =9.70	NDI Score	Neck Pain/ symptoms			Upper limb Pain/symptoms			Patient Input (weight)	Patient Input (Instruction/ Performance)	Patient Input (Ability to Perform)	Patient Input (Time & Effort)
				Before Testing	Immediately After Testing	24 H After Testing	Before Testing	Immediately After Testing	24 H After Testing				
2	50/60	9	29%	3	7	6	2	6	4	7	1	8	7
4	50/60	13	27%	4	7	4	2	6	4	6	1	7	7
9	50/60	11	25%	4	7	5	2	6	4	6	1	7	7
12	52/60	16	35%	4	6	5	2	7	6	6	1	6	5
15	55/60	15	39%	4	5	4	2	7	6	5	1	6	6
18	55/60	14	42%	5	7	5	2	7	6	5	1	6	6
20	55/60	17	30%	4	6	4	3	7	5	5	1	5	4
23	50/60	12	50%	5	6	5	2	7	6	6	1	6	5
26	55/60	14	40%	4	6	4	3	7	6	5	1	6	5
28	60/60	20↑	19%	3	5	3	0	4	4	4	1	5	5
58	55/60	15	25%	4	5	4	2	6	5	5	1	6	6
59	52/60	14	27%	4	5	4	3	6	5	5	1	6	6
60	48/60	6	40%	5	6	5	3	6	5	8	1	8	7
61	50/60	7	42%	5	6	5	2	7	6	8	1	8	8
62	40/60	3	60%	5	8	7	3	7	7	9	1	9	9
63	30/60	0↓	70%	5	7	5	3	6	4	9	1	9	9
64	35/60	5	48%	5	6	5	2	6	6	8	1	8	8
65	45/60	9	52%	5	6	6	3	6	5	7	1	7	7
66	40/60	6	55%	5	6	5	3	6	5	8	1	8	7
67	35/60	4	59%	5	6	5	3	7	6	8	1	8	8
68	30/60	0↓	74%	4	6	5	3	5	3	9	1	9	9
69	45/60	9	43%	4	6	5	3	6	5	7	1	7	7
70	40/60	4	60%	5	7	6	3		6	8	1	8	7

SAMP: Single Arm Military Press, NDI: Neck and Disability Index, Admin Time: Administration Time (Description, demonstration, instruction and 30S Performance). Completion Time (Warm-Up Time & Administration Time). ↓=Minimum Score. ↑=Maximum Score.

Appendix 21

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

Telephone Checklist: Preliminary Screening (Phone Screening) English and Arabic

A list of patients who have had visits to the Rheumatology and Physical Therapy Medicine Department at Tanta Universal Teaching Hospital (Egypt) with a diagnosis indicating non-specific uncomplicated neck pain was obtained.

All potential patient participants were telephoned to answer the following questions as follows:

Hello, I am phoning from Tanta Universal Teaching Hospital. We are working with Rheumatology and Physical Therapy Medicine Department to develop and evaluate health related outcome measure on patients suffering from neck pain as well as healthy subjects.

Would you like to hear about it?

Recent research suggested that neck problems frequently associated with upper limb pain/disability. We want to develop a physical performance test (outcome measure) to evaluate the upper limb functional capacity and used in the assessment and during the management process of patients with neck pain. This will help us to identify and quantify any upper limb disability and cure it while treating neck pain, which will make neck pain patients to feel better and help them to cope better with normal daily activity.

So, we are recruiting people with neck pain and healthy subjects. If you are interested in being involved in our study, you will be requested to attend:

For patient participants: two testing sessions at Tanta Universal Teaching Hospital. In these sessions, you will be met by the study's assessor who will tell you more about the study, ask you to complete questionnaires and carry out face-to-face assessment. If the testing procedure would be suitable and beneficial, you will be requested to perform a physical performance test for 30 seconds. Each session may take up to 45 minutes.

For healthy subjects: one single assessment and testing session at Tanta Universal Teaching Hospital. In this session, you will be met by the study's assessor who will tell you more about the study, ask you to complete questionnaires. If the testing procedure would be suitable, you will be requested to perform a physical performance test for 30 seconds. This session may take up to 45 minutes.

Would you be happy to participate?

- ☐ Yes
- ☐ No

If yes, may I ask you a few questions about your neck? This will help me to determine whether this testing procedure is suitable and good for you. Any information you give me will be kept confidential.

Patient Participants:

2. Have you had your neck symptoms for longer than 2 weeks? (*in weeks) (Acute/sub-acute or chronic pain)*
   - Yes
   - No

8. Have you had any treatment to your neck in the last 3 months? *(What diagnosis? What the practitioner told you about your neck problem)*

9. Are you planning to see anyone else for treatment?
   - Yes
   - No

If yes, to postpone until after the testing.

10. Do you have any other health problems such as dizziness, double vision, speech, swallowing, LOC (loss of consciousness)?
    - Yes
    - No

(Exclude patient with any major health problems)

11. Are you able to get on/off bed without help?
    - Yes
    - No

12. Are you able to walk, drive or use public transport without help?
    - Yes
    - No

13. Are you able to come to the new Tanta Universal Teaching Hospital?
    - Yes
    - No

Healthy Subjects:

1. Have you ever injured your head?
   - Yes
   - No

2. Have you ever injured your neck?
   - Yes
   - No

3. Have you ever injured your upper limb (Shoulder/Arm/Hand)?
   - Yes
   - No

4. Over the past 3 months, have you had any neck and/or upper limb symptoms?
   - Yes
   - No

5. Are you having any problems with dizziness, double vision, speech, swallowing, LOC (loss of consciousness)? If so, please specify?

6. Do you have any general medical problems? If so, please specify?

Thank you for taking the time to answer those questions. I can tell you that at this stage it would seem that:

c. The testing procedure would be suitable and beneficial for you (Make Appointment)
d. The testing procedure is not suitable for you (exclude this patient from the study).

Making appointment (if a)

Can you come to Tanta Universal Teaching Hospital on (day) at (time) for face-to-face assessment, which be followed by the testing? *Negotiate an appointment time*.

Study information sheet

I would like to send you the study information sheet, where you will find more information about the study and the testing. Please read it and you may discuss your participation with your family and/or friends before attending the single face-to-face assessment and testing session: your address is…

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

استمارة تشخيص المريض

للمريض: المرحلة الاولى من الفحص

من خلال مكالمة هاتفية سوف يتم تعريف المريض بمضمون البحث والغرض منه وما سبب اختياره ضمن عينة البحث لتوافر بعض الشروط متمثلة في الم الرقبة، وعند موافقته على الاشتراك في عينة البحث يطلب منه الاجابة على الاسئلة التالية:

١ – هل لديك أي الم في الرقبة لمدة اطول من اسبوعين؟

لا
نعم

٢ – هل عولجت خلال الثلاث شهور الماضية؟

لا
نعم (ما نوع التشخيص، ما هو تخصص المشخص)

٣ – هل تخطط لرؤية متخصص اخر بغرض العلاج قريبا؟

لا
نعم (يرجى التأجيل بعد اجراء الاختبار)

٤ – هل لديك أي مشاكل صحية اخرى مثل الدوار او ازدواج الرؤية او فقدان الوعي؟

لا
نعم (استبعاد أي مريض لدية مشاكل اخرى كبيرة)

٥- هل انت قادر على الحضور الى مستشفى طنطا التعليمي بدون مساعدة؟

لا (استبعاد أي مريض لا يستطيع الحضور لأسباب صحية)
نعم

نشكر الشخص المتحدث معنا لاستقطاعه جزء من وقته للإجابة على الاسئلة.
مع تحديد موعد لحضوره الى المستشفى التعليمي الجديد بطنطا في يوم ..... الساعة .....
معرفة العنوان الخاص به لإرسال معلومات أكثر عن البحث وإجراءاته واسباب اختياره في عينة البحث ومناقشة المشاركة في هذا البحث مع العائلة والاصدقاء المقربين قبل اتخاذ القرار النهائي في المشاركة.
المرحلة الاولى من الفحص: للأصحاء

١. هل عانيت في الماضي من إصابات في الراس؟

٢. هل عانيت في الماضي من إصابات في الرقبة؟

٣. هل عانيت في الماضي من إصابات في الكتف او الذراع او اليد؟

٤. في الثلاث شهور الماضية، هل عانيت من الام في الرقبة او الكتف او الذراع او اليد؟

٥. هل تعاني من أي مشاكل مرتبطة بالدوار او زغللة العيون والرؤية المزدوجة او صعوبة في الكلام او تورمات وإغماءات؟ لو حدث هذا رجاء اعطيني تفاصيل؟

٦. هل تعاني من أي مشكله صحية وطبيه في العموم؟

نشكر الشخص المتحدث معنا لاستقطاعه جزء من وقته للإجابة على الاسئلة.
مع تحديد موعد لحضوره الى المستشفى التعليمي الجديد بطنطا في يوم ..... الساعة .....
معرفة العنوان الخاص به لإرسال معلومات أكثر عن البحث وإجراءاته واسباب اختياره في عينة البحث ومناقشة المشاركة في هذا البحث مع العائلة والاصدقاء المقربين قبل اتخاذ القرار النهائي في المشاركة.

Appendix 22

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

Patient Information Sheet (English and Arabic)

We wish to invite you to participate in a research study. In order to have a clearer understanding of this research context, please read the following information sheet and do not hesitate to ask if there is anything that it not clear or you would like further information before you decide to take part.

What is the purpose of this study?

Neck pain is common, painful and many people report having trouble using their arm. Physiotherapists should try to measure the problems that people have when using their arm, so that they can advise patients how best to improve use of their arm.

The aim of this study is to develop and evaluate a physical performance test/performance-based outcome measure, the Single Arm Military Press (SAMP) Test, which is easy to use, economical, quick (maximum 2 minutes to perform and score). Preliminary research suggested that SAMP test is promising upper limb outcome measure for neck pain patients.

Why I have been invited?

We are inviting patients, age 18 years or over, with non-specific neck pain as well as healthy subjects to take part in this study. The patients we are looking for should be able to travel to the Rheumatology and Physical Therapy department/clinic at Tanta Universal Teaching Hospital without support.

You have been invited to participate in this research program because of the type of neck pain you suffering from or not suffering from any pain at all. If you are currently having treatment for your neck pain, had treatment in the past 3-month or have never need treatment for your neck or upper limb then you may be eligible for this study.

What will happen if I decide to take part?

If you are interested in being involved in this study:

Patient participants:

You will be asked to attend two testing sessions, in which each may take up to 45 minutes. You will be requested to complete two baseline questionnaires, which will give us more information about your neck and/or upper limb symptoms as well as your general physical and psychological well-being. This will be followed by face-to-face assessment to ensure that the testing procedure is suitable and beneficial for you. You will be then requested to sign a consent form (session 1 only) to say that you agree to be involved with this study and complete the testing for the first session.

Healthy subjects:

You will be asked to attend one single assessment and testing session, which may take up to 45 minutes. You will be requested to complete two baseline questionnaires, which will give us more information about your any neck and/or upper limb symptoms as well as your general physical and psychological well-being. This will be followed by face-to-face

assessment to ensure that the testing procedure is suitable for you. You will be then requested to sign a consent form to say that you agree to be involved with this study and complete the testing procedure.

It is preferable if you can wear a suitable, sleeveless/half-sleeves top, during the session so that the shoulder and elbow joints can be observed.

What do I have to do?
You will receive a complete demonstration and instruction of the test followed by warm-up and the test procedure, which will take up to 2-munites under direct supervision of a physiotherapist/physician.

Do I have to participate in this study?

'Only if you want to'

Participation is voluntary, you do not have to participate or you may withdraw from the study at any time before attending the face-to-face assessment and testing. However, please let us know if you are unable to participate at least 24-hours before your appointment. You do not need to tell us why you do not want to participate.

Are there any risks involved?
There are no known risks. This research program is simply validating performance-based outcome measure.

You may experience some muscle soreness, which is completely normal following physical exercise and may last up to 72 hours.

Confidentiality
All information from this study will be kept entirely confidential. All consent forms and any other identifiable information will be destroyed once the study has been completed. You will be informed of the results of the research if you wish.

Sheffield Hallam University | Centre for Health and Social Care Research

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

معلومات شاملة لإجراءات البحث

نتشرف بدعوة سيادتكم بالمشاركة كجزء من عينة هذا البحث ولفهم ومعرفه المعلومات الشاملة لإجراءات البحث برجاء قراءة المعلومات التالية وعدم التردد في التساؤل عما إذا كان هناك أي معلومة غير واضحة او مزيد من المعلومات قبل ان تقرر المشاركة.

ما هو الغرض من هذه الدراسة؟

ان الم الرقبة من الامراض شائعة الحدوث والتي تؤدى الى الم شديد في بعض الاحيان، ومعظم مصابي الم الرقبة يجدون صعوبة في استخدام الذراع والكتف وبالأخص الاعمال التي تؤدى أعلى مستوى الرأس لذا يتوجب على الطبيب المتخصص في علاج الم الرقبة باختبار مدى الاعاقة في الكتف والذراع واليد.

الغرض من اجراء هذه الدراسة هو تطوير طريقة قياس الاعاقة في الكتف والذراع واليد عند مرضى الم الرقبة.

بعض الابحاث العلمية الحديثة اثبتت ان طريقة القياس (SAMP) هي طريقة مثالية للقياس.

لماذا تم اختياري في عينة البحث؟

لقد تم اختيارك كجزء من عينة البحث لأنك عمرك أكثر من ١٨ سنة وتعاني من الم في الرقبة.

ماذا سيحدث إذا قررت المشاركة في هذا البحث؟

إذا وافقت على المشاركة في هذه الدراسة فستدعى للحضور مرة واحدة في توقيت يتراوح بين ٣٠: ٤٥ دقيقة.

سوف يطلب منك الاجابة على بعض الاسئلة الخاصة بألم الرقبة وحالاتك الصحية العامة وملء بعض الاستبيانات.

ملحوظة: ان ترتدى المريضة تي شرت لرؤية الذراع كاملة.

ماذا على ان افعل في هذا الاختبار؟

سوف تتلقى تعليمات واضحة متبوعة بنموذج للاختبار يليه الاحماء ثم اداء الاختبار تحت اشراف متخصص.

هل من الضروري المشاركة في هذه الدراسة؟
(إذا اردت انت المشاركة)

المشاركة في هذه الدراسة تطوعية ويمكنك الانسحاب حينما تشاء.

هل هناك أي مخاطر يمكن ان تنتج عن المشاركة في هذه الدراسة؟

لا توجد مخاطر على الاطلاق ولكن قد تشعر ببعض الارهاق العضلي البسيط والذي هو مشابهة للذي يحدث بعد ممارسة الانشطة البدنية وقد يستغرق ذلك ٧٢ ساعة للاستشفاء.

الخصوصية:

جميع المعلومات الخاصة بالمرضى في هذه الدراسة هي سرية تامة وسوف يتم تدمير جميع اوراق الموافقة واي معلومات اخرى عن المرضى بمجرد الانتهاء من الدراسة.

Appendix 23

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

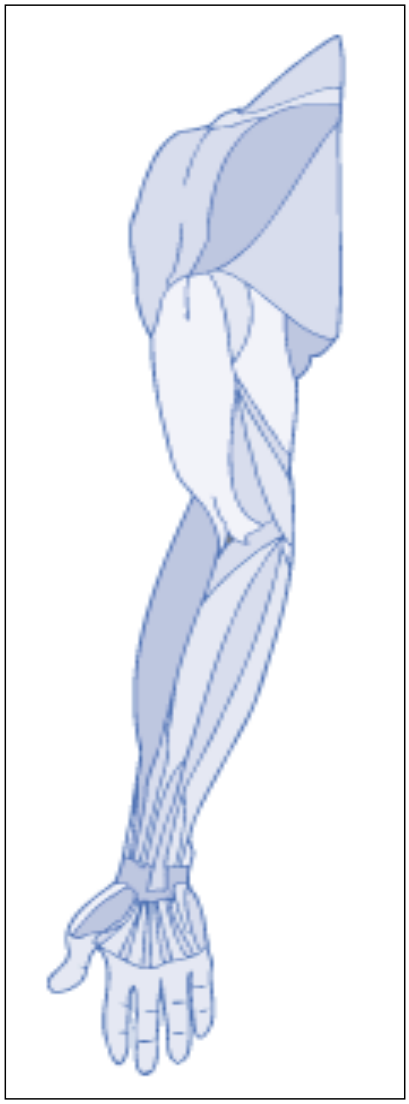
Visual Analogue Scale (VAS) 0-10 SAMP Reliability and Validity

Severity	Description of Experience
10 - Worst Possible Pain	I am in bed and can't move due to my pain. I need someone to take me to emergency room to get help for my pain.
9 – Extremely Sever Pain	My Pain is all that I can think about. I can barely talk or move because of the pain.
8 – Very Severe Pain	My pain is so severe that it is hard to think of anything else. Talking and listening are difficult.
7 – Severe Pain	I am in pain all the time. It keeps me from doing most activities.
6 – Distressing Pain	I think about my pain all the time. I give up many activities because of my pain.
5 – Distracting Pain	I think about my pain most of the time. I cannot do some of the activities I need to do each day because of the pain.
4 – Moderate pain	I am constantly aware of my pain but I can continue most activities.
3 – Uncomfortable Pain	My pain bothers me but I can ignore it most of the time.
2 – Mild Pain	I have a low level of pain. I am aware of my pain only when I pay attention to it.
1 – Slight Pain	My pain is hardly noticeable.
0 – No Pain	I have no pain

Appendix 24

DASH questionnaires Arabic Version

إعاقات الذراع والكتف واليد [1]



تعليمات

تستفسر هذه السلسلة من الأسئلة عن الحالات / الأعراض التي تحس بها في ذراعك، أو كتفك، أو يدك وعن مقدرتك على تأدية نشاطات معينة. الرجاء أن تجيب على *كل* سؤال، بناءً على حالتك خلال الأسبوع الماضي،
و ذلك بوضع دائرة حول الرقم المناسب.
إذا لم تسنح لك الفرصة لتأدية نشاط ما خلال الأسبوع الماضي، فالرجاء أن تقدّر بأفضل ما تستطيع لتختار الجواب الأقرب إلى الدقة. عند إجابتك علي الأسئلة، ليس مهماً أي يد أو ذراع تستخدم لتمارس نشاطك سواء كانت اليد المصابة أو السليمة.
الرجاء أن تجيب بناءً على مقدرتك بغض النظر عن الطريقة التي تؤدي بها العمل.

2

إعاقات الذراع والكتف واليد

الرجاء أن تقيّم قدرتك على فعل النشاطات التالية خلال الأسبوع الماضي، و ذلك بوضع دائرة حول الرقم الذي يقع تحت الجواب المناسب.

	بلا صعوبة	بصعوبة خفيفة	بصعوبة متوسطة	بصعوبة شديدة	غير قادر
1. أن تفتح علبة جديدة أو مُحكمة الإغلاق.	1	2	3	4	5
2. أن تكتب.	1	2	3	4	5
3. أن تبرم/ تدير مفتاحا (مثل أن تدير مفتاح السيارة لتشغيلها).	1	2	3	4	5
4. أن تُحضّر/ تعد وجبة طعام .	1	2	3	4	5
5. أن تدفع لتفتح بابا ثقيلا.	1	2	3	4	5
6. أن تضع شيئاً ما على رف فوق مستوى رأسك.	1	2	3	4	5
7. أن تقوم بأعمال المنزل الثقيلة (مثل غسل الحيطان أو إزاحة الأثاث أو سواها من الأشياء الثقيلة).	1	2	3	4	5
8. أن تعمل في الحديقة أو في فناء الدار.	1	2	3	4	5
9. أن ترتب السرير.	1	2	3	4	5
10. أن تحمل كيس التسوق أو حقيبة الوثائق.	1	2	3	4	5
11. أن تحمل غرضاً ثقيلاً (يزيد وزنه عن عشرة أرطال، أو أربعة كيلوغرامات و نصف).	1	2	3	4	5
12. أن تغيّر لمبة المصباح من فوق رأسك.	1	2	3	4	5
13. أن تغسل شعرك أو تنشفه بالمجفف الهوائي.	1	2	3	4	5
14. أن تغسل ظهرك.	1	2	3	4	
15. أن تلبس كنزه/ثوب/بلوزه (سترة ذات أكمام طويلة).	1	2	3	4	5
16. أن تستخدم سكيناً لتقطيع الطعام.	1	2	3	4	5
17. أن تقوم بنشاطات ترفيهية تتطلب جهداً خفيفاً (مثل لعب الشطرنج أو سواها من الألعاب الأخرى).	1	2	3	4	5
18. أن تقوم بنشاطات ترفيهية تبذل فيها بعض القوة أو الدفع عبر ذراعك أو كتفك أو يدك (مثل لعب التنس أو سواها من الألعاب الأخرى).	1	2	3	4	5
19. أن تقوم بنشاطات ترفيهية تحرك فيها ذراعك بحرية (مثل لعب رمي القرص أو الفريسبي أو سواهما من ألعاب مماثلة).	1	2	3	4	5
20. أن تتنقل بالمواصلات من مكان لآخر (أن تتنقل بمساعدة أعضاء جسدك العلويه كالإمساك بمقود السيارة).	1	2	3	4	5
21. النشاطات الجنسية. (الإجابه على هذا السؤال اختياري)	1	2	3	4	5

3

إعاقات الذراع والكتف واليد

	لا أبدا على الإطلاق	بشكل طفيف	بشكل متوسط	كثيراً	بشكل بالغ للغاية
22. خلال الأسبوع الماضي، هل أثرت المشكلة في ذراعك أو كتفك أو يدك بنشاطاتك الاجتماعية العادية مع عائلتك، أو أصدقائك، أو جيرانك، أو زملائك بالمهنة/النادي الاجتماعي؟ (ضع دائرة حول الرقم المناسب)	1	2	3	4	5

	غير محدود على الإطلاق	محدود بشكل طفيف	محدود بشكل متوسط	محدود جداً	غير قادر
23. خلال الأسبوع الماضي، هل أثرت المشكلة في ذراعك أو كتفك أو يدك بنشاط عملك أو أي نشاطات يومية اعتيادية أخرى؟ (ضع دائرة حول الرقم المناسب)	1	2	3	4	5

الرجاء تقدير شدة العوارض التالية التي أحسست بها خلال الأسبوع الماضي (ضع دائرة حول الرقم المناسب).

	لا يوجد	قليلاً	بشكل متوسط	بشدة	بشدة بالغة للغاية
24. وجع/ ألم/عوار في الذراع، أو الكتف، أو اليد.	1	2	3	4	5
25. وجع/ ألم/ عوار في الذراع، أو الكتف، أو اليد حينما أديت أي نشاط مُعيّن.	1	2	3	4	5
26. وخز (مثل وخز الدبابيس و الإبر) في ذراعك، أو كتفك، أو يدك.	1	2	3	4	5
27. ضعف في ذراعك، أو كتفك، أو يدك.	1	2	3	4	5
28. تيبّس/ تصلب في ذراعك، أو كتفك، أو يدك.	1	2	3	4	5

	لا صعوبة	صعوبة خفيفة	صعوبة متوسطة	صعوبة شديدة	صعوبة بالغة الشدة بحيث لا أقدر على النوم
29. خلال الأسبوع الماضي، كم كانت صعوبة نومك بسبب الوجع/ألم/عوار في ذراعك، أو كتفك، أو يدك؟ (ضع دائرة حول الرقم المناسب)	1	2	3	4	5

	لا أوافق بشدة	لا أوافق	لستُ موافقاً ولا مُعترضاً	أوافق	أوافق بشدة
30. أشعر بأني أقلّ ثقة بنفسي وذلك بسبب مشكلة ذراعي، أو كتفي، أو يدي (ضع دائرة حول الرقم المناسب).	1	2	3	4	5

إعاقات الذراع والكتف واليد: إجمالي درجات الإعاقات / الأعراض = [(مجموع عدد الإجابات) − 1] × 25 ، حيث (العدد) يساوي عدد الإجابات المكتملة.
العدد

لا يمكن حساب إجمالي الدرجات في مقياس إعاقات الذراع والكتف واليد إذا تجاوز عدد البنود الناقصة ثلاثة بنود.

4

إعاقات الذراع والكتف واليد

وحدة قياس العمل (اختياري)

الأسئلة التالية تستفسر عن تأثير مشكلة ذراعك، أو كتفك، أو يدك على مقدرتك على العمل (بما فيه القيام بالأعمال المنزلية إن كان ذلك هو دور عملك الرئيسي).
الرجاء أن تذكر ما هو عملك/ وظيفتك: _____
☐ أنا لا أعمل (يمكنك ترك هذا القسم).

الرجاء أن تضع دائرة حول الرقم الأفضل وصفاً لمقدرتك الجسدية خلال الأسبوع الماضي، هل عانيت أية صعوبة في

	لا صعوبة	صعوبة خفيفة	صعوبة متوسطة	صعوبة شديدة	غير قادر
1. أن تستخدم أسلوبك الاعتيادي في عملك؟	1	2	3	4	5
2. أن تؤدي عملك الاعتيادي، و ذلك بسبب وجع/ ألم/ عوار الذراع أو الكتف أو اليد؟	1	2	3	4	5
3. أن تؤدي عملك بشكل حسن مثلما تريد؟	1	2	3	4	5
4. أن تقضي نفس القدر من الوقت الذي تستغرقه عادةً لأداء عملك؟	1	2	3	4	5

وحدة قياس الرياضات / فنون الأداء (اختياري)

تتعلق الأسئلة التالية بتأثير مشكلة ذراعك، أو كتفك، أو يدك، على العزف على آلتك الموسيقية أو على لعب الرياضة أو كليهما.
إذا كنت تمارس أكثر من رياضة، أو تعزف على أكثر من آلة موسيقية، (أو الاثنين معاً)، فالرجاء الإجابة بالنظر إلى ذلك النشاط الذي تعتبره الأهم بالنسبة إليك.
الرجاء الاشارة إلى الرياضة أو الآلة الموسيقية الأكثر أهمية بالنسبة لك: _____
☐ أنا لا ألعب أي رياضة أو أعزف على أي آلة موسيقية (يمكنك ترك هذا القسم).

الرجاء أن تضع دائرة حول الرقم الأفضل وصفاً لمقدرتك الجسدية خلال الأسبوع الماضي، هل عانيت أية صعوبة في

	لا توجد صعوبة	صعوبة خفيفة	صعوبة متوسطة	صعوبة شديدة	غير قادر
1. أن تستخدم أسلوبك الاعتيادي في عزفك على آلتك الموسيقية أو لعبك لرياضتك؟	1	2	3	4	5
2. أن تعزف على آلتك الموسيقية، أو تلعب الرياضة التي تحبها بسبب وجع/ ألم/ عوار الذراع أو الكتف أو اليد ؟	1	2	3	4	5
3. أن تعزف على آلتك الموسيقية أو تلعب رياضتك بشكل جيد مثلما تحب؟	1	2	3	4	5
4. أن تقضي نفس الوقت الاعتيادي في التمرين أو في العزف على آلتك الموسيقية أو لعب رياضتك؟	1	2	3	4	5

للحصول على إجمالي درجات وحدات القياس الاختيارية: اجمع القيم الرقمية لكل جواب، و قسّمها على الرقم 4 (عدد البنود)، ثم اطرح منها العدد 1، و من ثم اضرب الحاصل بالرقم 25. لا يمكن حساب إجمالي درجات وحدة القياس الاختيارية إذا كانت هناك أية بنود ناقصة.

Appendix 25

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

Face to face assessment: (English and Arabic)

Participant's demographic information

Assessment date	--/--/----
Participant name	
Participant ID number (in the study)	
Date of Birth	--/--/----
Occupation	
Weight	
Height	
Telephone number	
Email address	
Home address	

- Where is your symptoms now? *Complete the body chart*.

- How severe is your symptoms now?

  Neck pain/symptoms
  No pain 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 worst possible pain (*see Appendix 11*)

  Upper limb pain/symptoms
  No pain 0 – 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9 – 10 worst possible pain (*see Appendix 11*)

- Over the past 2 weeks are your symptoms:
  - Getting better
  - Getting worse
  - Same

If worsening, in what way? (exclude deteriorating neurological condition e.g. cord sign, radiculopathy)

- Were you involved in an accident which caused your pain? (exclude recent major trauma)

- How long have you had your neck symptoms? In weeks 'determine whether acute, sub-acute or chronic'

- How long ago did you first experience these symptoms? In weeks 'determine recurrence' and how it was treated?

- Have you had any treatment such as physical therapy in the past three months?

- Have you ever injured your shoulder, arm or hand substantially? exclude injuries which has resulted in current or prolonged disability
- Is your weight steady? Exclude unintentional weight loss

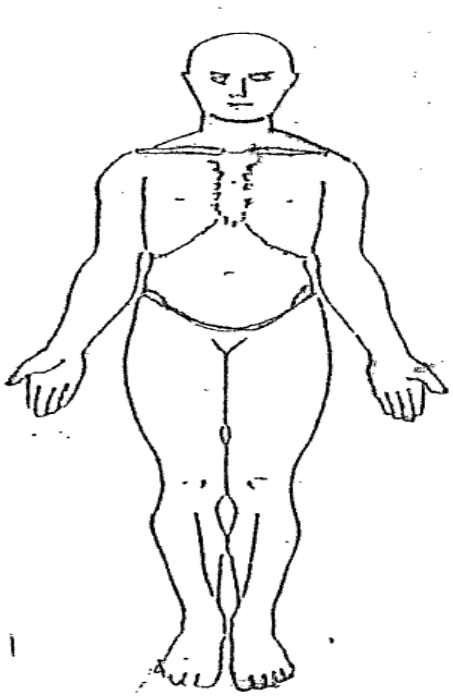
- Are you sleeping OK at night? Exclude severe pain at night

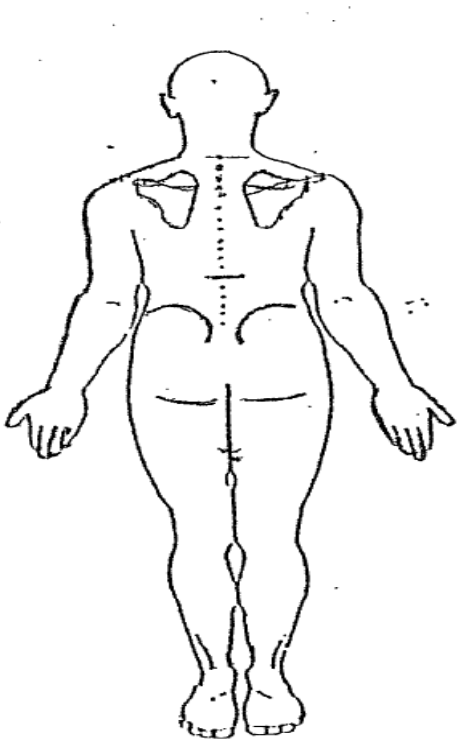
- Are you having any problems with dizziness, double vision, speech, swallowing, LOC (loss of consciousness)? Exclude in accordance with the criteria

- Do you have any general medical problems? If so, specify what type of problems? (Exclude vertebral artery problems)

- Do you have any general medical problems? If so specify, (Exclude – severe rheumatoid arthritis, severe multiple sclerosis, cancer, osteoporosis, cardiac conditions, severe SOBOE, uncontrolled hypertension, postural hypotension, balance problems).

If all answers confirmed eligibility and the NDI scored at least 10%, and patient still happy to proceed. Patient should sign the consent form and allocated for SAMP testing.





Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

المرحلة الثانية من الفحص (بحضور المريض):

تاريخ يوم الفحص	...../...../.....
اسم المريض	
رقم المريض	
تاريخ ميلاد المريض	
وظيفة المريض	
الوزن	
الطول	
رقم تليفون المريض	
عنوان المريض	
مرحلة البحث	

السؤال الاول:
اين يوجد الالم الان؟ (استكمال مخطط الجسم المرفق)

السؤال الثاني:
ما مدى قوة الالم؟

الم الرقبة:
لا يوجد الم 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 اقصى الم ممكن

الم الذراع والكتف واليد:
لا يوجد الم 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 اقصى الم ممكن

السؤال الثالث:
خلال الاسبوعين الماضيين الم الرقبة:
تحسن:
ازداد سوء: (استبعاد الحالات المرتبطة بالعمود الفقري والاعصاب)
نفس الألم:

السؤال الرابع:
هل الم الرقبة الذي تشعر به نتيجة لحادث؟
لا:
نعم: (استبعاد المرضى المصابون حديثا)

السؤال الخامس:
ما مدة شعورك بألم الرقبة في الوقت الراهن؟ (بالأسبوع)

السؤال السادس:
متى شعرت بألم الرقبة للمرة الاولى؟ (بالأسبوع)
.....

السؤال السابع:
هل تلقيت أي علاج في الثلاث شهور الماضية؟
لا:
نعم:

السؤال الثامن:
هل سبق لك ان تعرضت لإصابة في الكتف او الذراع او اليد؟
لا:
نعم: (استبعاد المصابون بالألم شديد او إعاقة في الكتف او الذراع او اليد)

السؤال التاسع:
هل وزنك ثابت؟
لا: (استبعاد أي نقصان في الوزن بدون قصد)
نعم:

السؤال العاشر:
هل تنام ليلا بدون اضرابات؟
نعم:
لا: (استبعاد الالم الليلي)

السؤال الحادي عشر:
هل لديك أي مشاكل صحية مثل الدوار، ازدواج الرؤية، او فقدان الوعي؟
لا
نعم: (استبعاد المرضى مع هذه المشكلات الطبية)

إذا كانت الاجابات مطابقة لأسس اختيار المريض ونتيجة مقياس إعاقة الرقبة (NDI) ١٠ ٪ على الاقل يتم توقيع المريض على موافقته للاشتراك في عينة البحث واجراءه.

Appendix 26

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

Study Consent Form (English and Arabic)

Name of the researcher: Ahmad Alreni

1. I confirm that I have read and understand the information sheet dated 0 /0 /00 for the above research study and have had opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason. ☐

3. I agree to tack part in the above research study ☐

Name of the subject (BLOCK CAPITALS):

……

Signature:

……Date:

……

Signature of the researcher:

……

Sheffield Hallam University | Centre for Health and Social Care Research

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

موافقة المريض على الاشتراك في عينه البحث

اسم الباحث: احمد صلاح الدين العريني

١. اقر انا الموقع ادناه على إنني قرأت وفهمت المعلومات الخاصة بإجراءات هذا البحث وكان لدى المقدرة على الاستفسار عن بعض الاجزاء الغامضة لدى في اجراءات البحث وتم الاجابة على استفساراتي

٢. اتفهم انا الموقع ادناه إنني اشارك تطوعيا في هذا البحث مع مقدرتي في الانسحاب منه بدون ابداء أي اسباب

٣. اوافق على الاشتراك في عينة هذا البحث

اسم المريض: .....
التوقيع: .....
التاريخ: .....

Appendix 27

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

Data Collection Sheet (English and Arabic)

Assessment Date:
Examiner Name:
Group Number:
Weight used:

Patient ID Number	SAMP Score reps/30-sec	Administration Time	Completion Time

SAMP score = number of repetitions in the 30-second, **Administration time** = description, demonstration, instructions and the 30-sec performance, **Completion time** = warm-up time and administration time.

Sheffield Hallam University | Centre for Health and Social Care Research

Measuring upper limb disability in neck pain population: evaluation of the reliability and validity of the Single Arm Military Press (SAMP) test

استمارة جمع البيانات

تاريخ اداء الاختبار: .....
اسم الفاحص: .....
رقم المجموعة: .....
الوزن المستخدم: .....

رقم المريض	ت \ ٣٠ث	زمن الاختبار	الزمن الكلى الاختبار

- ت\ ٣٠ ث = عدد التكرارات على ٣٠ ثانية
- زمن الاختبار = شرح الاختبار، اداء نموذج، التعليمات + ٣٠ ث اداء الاختبار
- الزمن الكلى للاختبار = الاحماء + زمن الاختبار