Neck Pain: “a pain in the neck”?  
A study of therapeutic modalities and clinimetrics
The study presented in this thesis was performed at the EMGO-institute of the VU University Medical Center Amsterdam, the Netherlands. The EMGO-institute participates in The Netherlands School of Primary Care Research (CaRe), which was re-acknowledge in 2000 by the Royal Netherlands Academy of Arts and Science (KNAW).

The study presented in this thesis was supported by the Netherlands Organisation for Health Research and Development (ZonMW), grant no-940-31-060.

Financial support for publication of this thesis has been kindly provided by the EMGO-institute VUmc, the Dutch Association of Manual Therapy (NVMT), the Royal Dutch Association for Physical Therapy (KNGF), SOMT Master Education in Manual Therapy Amersfoort, Primary Health Care Organisation Zoetermeer (SGZ) and Medical Center IMPACT Zoetermeer.

ISBN

Cover-design: Wim Groeneveld
Printed by: Gildeprint Drukkerijen B.V. Enschede

© 2007, Jan Pool

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system without prior written permission in writing from the author.
Neck Pain: “a pain in the neck”? 
A study of therapeutic modalities and clinimetrics

ACADEMISCH PROEFSCHRIJT

ter verkrijging van de graad Doctor aan de Vrije Universiteit Amsterdam, op gezag van de rector magnificus prof.dr. L.M. Bouter, in het openbaar te verdedigen ten overstaan van de promotiecommissie van de faculteit der Geneeskunde op woensdag 20 juni 2007 om 15.45 uur in de aula van de universiteit, De Boelelaan 1105

door

Johannes Jacobus Maria Pool

geboren te Voorburg
promotoren: prof.dr.ir. H.C.W. de Vet
          prof.dr. L.M. Bouter
copromotor: dr. R.W.J.G. Ostelo
Voor Annelies

Voor Sabrina, Stephanie, Lisanne en Jesper
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1</td>
<td>Introduction</td>
<td>9</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>A recent update of conservative treatment and cost-effectiveness for acute and chronic neck pain</td>
<td>17</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Comparison of the effectiveness of a behavioural graded activity program and manual therapy in patients with sub-acute neck pain: design of a randomised clinical trial</td>
<td>37</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Is a behavioural graded activity program effective in patients with sub-acute neck pain? a randomised clinical trial</td>
<td>61</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain?</td>
<td>85</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>The inter-examiner reproducibility of physical examination of the cervical spine</td>
<td>107</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>Reproducibility of cervical range of motion in patients with neck pain</td>
<td>125</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>The applicability of the Tampa Scale of Kinesiophobia in sub-acute neck pain patients: a qualitative study</td>
<td>143</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Added value of qualitative studies in the development of Health Related Patient Reported Outcomes</td>
<td>157</td>
</tr>
<tr>
<td>Chapter 10</td>
<td>Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain</td>
<td>173</td>
</tr>
<tr>
<td>Chapter 11</td>
<td>General Discussion</td>
<td>187</td>
</tr>
<tr>
<td>Contents</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>Samenvatting</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>Dankwoord</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td>List of publications</td>
<td>231</td>
<td></td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>235</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The demand for evidence based medicine is growing. Clinicians as well as policymakers and insurance companies want to be informed about the best available evidence for the effectiveness of treatment modalities. Furthermore, evidence is needed about clinimetrical properties of frequently used clinical tests or questionnaires, to diagnose or to evaluate the clinical status of a patient. Considering neck pain, a common musculoskeletal disorder, there still is a lack of evidence in favour of any treatment modality, specific clinical test or questionnaire. Moreover, to facilitate clinical decision making and reasoning concerning the choice of treatment, information regarding prognostic factors is of importance but the outcome of studies about these factors is inconclusive.

Is neck pain a “pain in the neck” for clinicians? To answer this question sufficiently we studied all mentioned factors; treatment modalities, some clinimetrical properties of clinical tests and questionnaires, and prognostic factors.

In this chapter we will briefly introduce the main research questions of this thesis.

Epidemiology

Neck pain is a common musculoskeletal disorder. The main feature of neck pain is pain in the cervical region which is often accompanied by other complaints as restriction of the range of motion and/or functional limitations. Often the neck pain is precipitated or aggravated by neck movements or sustained positions. Besides pain and stiffness, symptoms as for example, headache, brachialgia or dizziness may be present. There is no conclusive evidence regarding specific pathology in the majority of cases for acute or chronic neck pain. The pain can originate from many structures in the cervical region. Consequently, most cases are labelled as non-specific neck pain or neck pain of unknown origin. In this thesis, neck pain is defined as pain and disability in the cervical region, with or without radiating symptoms in the arms or head. Estimations indicate that 67% of individuals will suffer neck pain at any stage of life.
The point prevalence for non-specific neck pain in the general population varies between 9% and 22%\textsuperscript{3;21}. Woman are more likely to have neck pain than man (OR 1.8). Approximately one-third of all adults experience neck pain during the course of a year\textsuperscript{6}. In most cases neck pain is benign and self-limiting. However, in 5-10% of patients with non-specific neck pain will develop to a chronic pain disorder\textsuperscript{3}.

40.8% of individuals who suffer neck pain will contact their General Practitioner (GP) annually\textsuperscript{21}. The point prevalence in general practice is approximately 22 per 1000 registered patients annually. The management of neck pain patients, by the GP, comprises of information regarding the course of the complaints and a “wait and see policy”, sometimes accompanied by a prescription of analgesic medication or muscle relaxants. In addition, in one third of the patients the GP will also refer patients to the a physical and/or manual therapist\textsuperscript{21}.

### Interventions

In managing neck pain patients a physical and/or manual therapist can choose between several therapeutic interventions, depending on the aim of the treatment. However, the question arises which intervention will be the most appropriate. Reviews performed by the Cochrane Collaboration\textsuperscript{9;10;14;16} and others come to the same conclusion: the effectiveness of many commonly used conservative interventions for neck pain is still unclear. Manipulation and/or mobilization when used in combination with exercises seems the most promising intervention.

Hoving et al\textsuperscript{11;13} conducted a randomised clinical trial (RCT), including an economic evaluation\textsuperscript{15}, which was designed to examine the (cost)-effectiveness of manual therapy, physical therapy, and standard general practice. This study demonstrates that manual therapy is more effective than physical therapy or standard medical care by the GP for patients with neck pain. The total costs of manual therapy were approximately one-third the cost of the physical therapy and standard medical care by the GP. In this trial 49% of the population had sub-acute non-specific neck pain, while sub-acute is defined as neck pain between 4 and 12 weeks.

Sub-acute non-specific neck pain patients form an interesting subgroup since from 4 to 12 weeks a transition from acute to chronic neck complaints takes place, if chronic pain is defined by its duration. One possible explanation for this transition could be the role of psychological and social factors in the awareness of pain\textsuperscript{7;8;17}. Some sub-acute neck pain patients become enmeshed in a downward spiral of increasing
avoidance behaviour, disability and pain. Specially patients who interpret pain as threatening (pain catastrophizing) can exhibit kinesiophobia or fear of movement\textsuperscript{19,24}. Consequently, to prevent chronicity, the intervention should be more focussed on psychological and social factors. With this in mind, the emphasis of some interventions is placed on behavioural treatment, based on operant, cognitive or respondent techniques, influencing these psychological and social factors\textsuperscript{20,22-24}. Behavioural treatment focuses on reducing disability through the modification of environmental contingencies and cognitive processes. However, knowledge about the effectiveness of a behavioural intervention in sub-acute neck pain patients is not known. In this thesis we compared a behavioural graded activity programme with manual therapy, the most effective therapy for neck pain so far.

**Risk factors and prognostic factors**

Risk factors for neck pain are well documented, e.g. physical load as vibration, flexion of the neck, sitting posture and heavy lifting\textsuperscript{1}. Furthermore, high pain intensity and a previous history of neck pain are strongly associated with an unfavourable prognosis\textsuperscript{6,13}. Psychological factors, such as passive coping, cognition, fear avoidance, depression, anxiety, and social factors have also been reported to aggravate and perpetuate neck pain and found to be influencing the course of the complaints\textsuperscript{1,4,12,17,18}. However these studies which have been carried out, were heterogeneous with regard to their study population and outcome measures. It is likely that patients with non-specific neck pain comprise several subgroups with different causes and different prognostic profiles. Information regarding prognostic factors, can facilitate clinical reasoning as well. Knowledge of prognostic factors which can identify subgroups at risk for poor outcome will influence the clinicians choice of treatment. Therefore, in this thesis, prognostic factors in neck pain patients are studied as well. We focussed in this study on psychological prognostic factors in non specific sub-acute neck pain patients.

**Clinimetrics**

To get insight into a patient’s representation of cervical complaints it is essential to use valid and reproducible instruments during clinical assessment of a patient to increase insight in function, activities in daily life, pain level, coping style, cognitions. Valid, reproducible and
responsive instruments are also a necessity to evaluate interventions. Many tests are available; mobility tests, segmental mobility tests, muscle tests, neuro-dynamic tests, neurological tests etc, but knowledge of their clinimetric properties, is still lacking. Besides functional tests, also questionnaires are used to evaluate the clinical status of the neck pain patient. Questionnaires such as the Neck Disability Index, the Numerical Rating Scale for pain and some psychological questionnaires like the Tampa scale of Kinesiophobia and the Pain Coping and Cognition List. The knowledge of their clinimetric properties, such as validity, reproducibility or responsiveness, is scarce or not known for the population neck pain patients.

Within this thesis quantitative research has been carried out to assess the reproducibility and discuss the outcomes of several clinical tests and questionnaires for neck pain. Furthermore, we also carried out a qualitative study. Reason for this line of research is the knowledge that enormous variations are present in the patient population with neck pain, for example demographic, ethnic as well as cultural differences which can influence the patient's attributions and beliefs. It is evident that, due to these differences, several items within a questionnaire can be interpreted differently. This is important knowledge for a clinician who will use the outcome of a questionnaire for e.g. clinical reasoning.

**General objectives are:**

1. What is the effectiveness of a behavioural graded activity programme in patients with sub acute neck pain compared to manual therapy?
2. Which psychological factors can influence the clinical course of neck pain?
3. Are clinical tests to assess mobility of the cervical spine reproducible.
4. Is the interpretation of item’s within a psychological questionnaire for patients with sub acute neck pain in line with the expectations of the researcher?
5. What is the Minimal Clinical Detectable Change score of questionnaires like the Neck Disability Index and the Numerical Rating Scale for pain, in patients with sub-acute neck pain?
Chapter 1

Outline of this thesis

This thesis is divided in two parts. The first part of this thesis concerns an overview of interventions and describes a randomised clinical trial. Chapter 2 reports the evidence of interventions used in daily practice. A description of the design of the RCT is reported in Chapter 3. The effectiveness of a Behavioural Graded Activity programme compared to manual therapy is assessed in Chapter 4. Furthermore we will discuss psychological prognostic factors who can influence the course of the neck pain in Chapter 5.

The second part concerns clinimetrical properties of frequently used clinical tests and questionnaires. Chapter 6 and 7 provide insight into the reproducibility of clinical tests (passive segmental and total cervical mobility) and a measurement instrument (digital inclinometer) used in daily practice of the physical or manual therapists. In chapter 8 and 9 the focus is placed on qualitative research of the Tampa Scale of Kinesiophobia, a scale measuring fear of movement/avoidance and the Pain Coping and Cognition List, a scale measuring cognitions, coping strategies and locus of control.

The minimal clinical important change of two questionnaires, the Numerical Rating Scale and the Neck Disability Index is discussed in Chapter 10.

This thesis will be concluded with a general discussion (Chapter 11) and a summary in both English and Dutch.
Reference list

Chapter 1


A recent update of conservative treatment and cost-effectiveness for acute and chronic neck pain

A part of this article is published as; Jan Pool, Sidney Rubinstein, Maurits van Tulder. Anerkannte Evidenz der Wirksamkeit konservativer Behandlungen akuter und chronischer Nackenschmerzen. Eine Übersicht. Manuelle Medizin 2005. 43:297-304

Submitted as; Jan J.M. Pool, Sidney M. Rubinstein, Frieke Vonk, Maurits van Tulder
A recent update of conservative treatment and cost-effectiveness for acute and chronic neck pain.
Abstract

The demand for evidence based medicine (EBM) has increased dramatically in the last decade. In this article, we discuss the benefit of EBM in general, and its role in the treatment of chronic neck pain, in specific. Although much evidence for conservative therapy for chronic neck pain is inconclusive, manipulative therapy and/or mobilization in combination with exercise seems to have the most promising results. Additionally, manipulative therapy would appear to be more cost-effective than physical therapy or standard medical care (as administered by the general practitioner).
Introduction

Neck pain

Neck pain is a common musculoskeletal disorder. The point prevalence of neck pain in the general population of the Netherlands varies between 9% and 22%⁴;²⁴, with approximately one-third of all adults experiencing neck pain during the course of a year⁶. Some 5-10% of these subjects will develop chronic pain⁴. The incidence of Whiplash Associated Disorders (WAD) varies in different countries, 106 per 100.000 in Australia and 94-188 per 100.000 in the Netherlands, for example, and there is much discussion in the literature about the natural course and epidemiology of WAD²⁶. Additionally, there is discussion in the literature whether mechanical neck disorders can be compared to WAD. For the purpose of this review, we distinguish WAD from mechanical neck pain.

The main feature of mechanical neck pain is pain in the cervical region, which is often accompanied by restriction of the range of motion and associated with functional limitations¹. The pain may originate from many structures in the cervical region, especially the spine and soft tissues, but there is no conclusive evidence regarding specific pathology in the majority of cases of acute or chronic mechanical neck pain². Consequently most cases are labeled as non-specific mechanical neck pain or mechanical neck pain of unknown origin². Risk factors for mechanical neck pain are physical load factors, such as vibration, flexion of the neck, sitting posture and heavy lifting¹. However, psychological factors, such as passive coping, cognition, fear avoidance, depression, anxiety and social factors are also reported to aggravate and perpetuate neck pain¹;¹⁹. High pain intensity and a previous history of neck pain are strongly and consistently associated with an unfavorable prognosis⁶;¹². Although mechanical neck pain is self-limiting, 40% contact their GP, while 30% of these patients are referred for further diagnosis by a medical specialist, and 32% are referred for conservative therapy, consisting of physiotherapy, manual therapy or chiropractic care⁴;²⁴. In order to encourage the use of evidence-based medicine for neck pain in clinical practice, we summarize the available evidence on the treatment for aspecific neck pain.

Evidence-based medicine

The importance of evidence-based medicine (EBM) has steadily increased during the past decade. EBM is defined by Sackett et al as
“Conscientious, explicit and judicious use of current best evidence in making decisions about care of individual patients”. The practice of EBM means integrating individual clinical expertise with the best available evidence derived from systematic reviews. However, ‘evidence’ is a rather broad concept. On the one hand, the evidence may refer to new or existing interventions, and these may be diagnostic, preventive and/or therapeutic. Evidence on the effectiveness of therapeutic interventions may be obtained through randomized clinical trials (RCTs), while evidence on the effectiveness of diagnostic interventions may be obtained through either RCTs or specific diagnostic studies. On the other hand, evidence on adverse effects or risk factors associated with a particular treatment are typically obtained from prospective, observational studies due to the lower incidence of adverse reactions. Furthermore, full economic evaluations provide evidence on cost-effectiveness and/or cost-utility.

While access to the internet has provided the clinician with a wealth of information, this has also resulted in information overload, therefore, hindering the clinician’s ability to synthesize the information regarding the management of neck pain. In EBM, information about the individual patient with his or her individual problem is collected from history taking, physical examination and additional diagnostic evaluation combined with clinical scientific information about diagnostic tools, prognostic factors and effectiveness of interventions. Sackett proposed five steps (See table 1) on how to practice EBM as a clinician.

**Table 1:** The 5-step model of EBP (Sackett)

<table>
<thead>
<tr>
<th>How to practice evidence-based medicine?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ask clinical questions you can answer</td>
</tr>
<tr>
<td>2. Search for the best evidence</td>
</tr>
<tr>
<td>3. Critically appraise the evidence</td>
</tr>
<tr>
<td>4. Apply the evidence in care for your patient</td>
</tr>
<tr>
<td>5. Self-evaluation (of the above steps)</td>
</tr>
</tbody>
</table>

Although we are more knowledgeable than years ago, the publication of more than 40,000 biomedical journals, 2,000,000 articles, and 20,000 books each year, has clearly led to an overload of information for the clinicians. The consequence is, they can no longer assimilate the best
available evidence. Systematic reviews have been conducted in order to resolve this problem. Within the field of therapeutic interventions, for example, systematic reviews pose a specific question, conduct a search strategy aimed at identifying relevant trials, and, conduct a critical appraisal of the methodological quality of the included trials. The result of this procedure is an unbiased and comprehensive view of the literature on that topic. Thus, for the clinician, a systematic review is an efficient manner of obtaining an answer to a clinically relevant question. To date, a substantial number of systematic reviews on neck pain have been published.

The Cochrane Collaboration

The Cochrane Collaboration is an international non-profit organization that prepares, maintains, and disseminates systematic up-to-date reviews of health care interventions. The main purpose of the Cochrane Collaboration is to provide information that is evidence-based, easily accessible, internationally developed, quality controlled, clinically useful, and periodically updated. One of the review groups, the Cochrane Back Review Group, coordinates reviews on back pain, neck pain and other spinal disorders. The Editorial Board of the Cochrane Back Review Group developed guidelines to facilitate a more systematic approach to the literature reviews, decrease the potential for bias, improve the quality of reviews in the field, facilitate comparison across reviews, and enhance consistency among reviewers. These systematic reviews have in turn served as the basis for a number of clinical guidelines for the primary care management of back and neck pain.

Methods

The Cochrane Library 2006, Issue 3, includes several reviews concerning various domains of neck pain. Two reviews of surgical or invasive interventions are not included in this present summary. One review included WAD and the others, mechanical neck disorders, with or without associated headache or radicular findings. Systematic reviews were included if they included randomised trials on acute (less than 6 weeks), sub acute (6-12 weeks), and/or chronic (more than 12 weeks) neck pain. For the evaluation of cost-effectiveness, Pubmed was searched for systematic reviews or economic evaluations using the free text words, ‘cost effectiveness’ and ‘economic evaluation’. Non-surgical treatment for neck pain was included and one
of the following outcome measures was required in order to be included in our analysis, pain, overall improvement or satisfaction with treatment, function (e.g. neck specific functional status), well-being (e.g. quality of life), disability (e.g. activities of daily living, work absenteeism) and side effects. The methodological quality assessment, data extraction and data analysis of the original systematic reviews were perused in this overview.

Cochrane reviews of neck pain

Whiplash Associated Disorders

In the review of Verhagen et al., WAD grade 1 and 2 was assessed which was defined as patients with neck complaints with or without musculoskeletal signs. 15 studies were included. Only one study was identified for chronic neck pain. The other studies included a heterogeneous study population of acute, sub acute and chronic neck pain. Several treatments were included in this review, which included immobilization with a soft collar, early active immobilization, pulsed electromagnetic therapy (PEMT) and multimodal treatment. In conclusion they found moderate evidence from several trials for the effectiveness of active treatment vs. passive treatment, and limited evidence from one trial for “act as usual”, multimodal treatment or PEMT.

In a review of Kay, moderate evidence was found for long-term benefit of exercise for WAD with or without headache. There is limited evidence of benefit for active range-of-motion exercises or a home exercise program for acute mechanical neck disorder including WAD. Kroeling et al found limited evidence for low and high frequency PEMF compared to placebo for acute WAD.

Mechanical neck disorders

Manipulation/mobilization

The review from Gross et al included 33 trials. This review found strong evidence for manipulation and/or mobilization when used in combination with exercises, although the type of exercises which were used was not mentioned in most of the studies. Manipulation and/or mobilization alone, however, were not found to be beneficial. Additionally, there is insufficient evidence for the effectiveness of manipulation and/or mobilization for radicular complaints.
Exercises
Kay et al\textsuperscript{15} found strong evidence for a multimodal care approach of exercise combined with manipulation or mobilizations to be beneficial for pain, function and global perceived effect for sub acute and chronic mechanical neck pain with or without headache. Moderate evidence was found for short and long term benefit of exercise, stretching and/or strengthening in chronic mechanical neck pain. Also moderate evidence was found for the short-term benefit of vertigo/ eye-fixation exercise imbedded in a more complex program for chronic mechanical neck disorders. Further, Kay et al found limited evidence of the benefit of strengthening exercise in the short and long term for chronic MND. The author couldn’t conclude what the relative benefit of exercise was compared to other treatments and what the relative benefit was of different exercise approaches.

Multidisciplinary biopsychosocial rehabilitation
Karjalainen et al. found limited evidence on multidisciplinary biopsychosocial rehabilitation for neck and shoulder pain\textsuperscript{14}. Only two relevant studies were included. There was little scientific evidence for the effectiveness of multidisciplinary biopsychosocial rehabilitation compared with other rehabilitation facilities for neck and shoulder pain. Another review on work hardening for workers with back and neck pain\textsuperscript{27} concluded that physical conditioning programs can be effective for back pain, especially if they include a cognitive behavioral approach. A subset of 19 trials were included in this review but none of these trials studied the effectiveness of interventions on workers with neck pain.

Physical modalities
Kroeling et al\textsuperscript{17} did not find convincing evidence of a clinically important benefit of electrotherapy, covering direct or modulated Galvanic current, Diadynamic current, iontophoresis, TENS, EMS, pulsed electromagnetic field (PEMF) and permanent magnets. They only found limited evidence for low and high frequency PEMF compared to placebo for chronic mechanical neck pain.

Massage
Haraldson et al\textsuperscript{11} included 19 trials in their review and assessed massage alone or massage in combination with other modalities. This review concluded that there was no significant advantage over no treatment, hot packs, exercises, sham laser, TENS, manual traction, mobilization, education or pain medication.
Medicinal and injection therapies
Peloso et al\textsuperscript{23} found that intramuscular injection of lidocaine for chronic mechanical neck disorders and an intravenous injection of methylprednisolone for acute whiplash patients were effective treatments. There was limited evidence for epidural injections of methylprednisolone and lidocaine for neck disorders with radicular findings. The effects of muscle relaxants and non-steroidal anti-inflammatory drugs are unclear. There was moderate evidence in favor of Botos-A intramuscular injection compared to saline.

Acupuncture
Trinh et al\textsuperscript{31} included 10 trials that examined acupuncture treatments on chronic neck pain. They concluded that there was moderate evidence that acupuncture relieves pain better than some sham treatments or waiting list controls. The effects are measured on pain and especially on the short term.
### Table 2. Reviews within framework of the Cochrane Collaboration

<table>
<thead>
<tr>
<th>Study</th>
<th>Studies included</th>
<th>Interventions</th>
<th>Evidence</th>
<th>Outcome</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verhagen 2004</td>
<td>15</td>
<td>Passive vs placebo/no treatment</td>
<td>Limited in favour of passive</td>
<td>Pain, GPE*</td>
<td>1 High quality study</td>
</tr>
<tr>
<td>WAD</td>
<td></td>
<td>Active vs no treatment</td>
<td>Limited in favour of active</td>
<td>Pain</td>
<td>2 High quality studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active vs passive</td>
<td>Conflicting evidence</td>
<td>Pain, ROM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active vs active</td>
<td>Conflicting evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Conflicting evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross 2004</td>
<td>33</td>
<td>Manip alone vs sham</td>
<td>Moderate evidence of no effect</td>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>MND 19, headache</td>
<td></td>
<td>Manip vs various comparisons</td>
<td>No evidence of benefit</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td>of cervical origin</td>
<td></td>
<td>Mob alone vs various comp.</td>
<td>Moderate evidence of no difference</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td>WAD 6, degenerative changes 6, radiculair findings 4</td>
<td></td>
<td>Manip and mob vs placebo</td>
<td>No evidence of benefit</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manip and mob vs no treatment</td>
<td>Short/long term benefit</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multimodal + physical agents</td>
<td>Moderate evidence of no effect</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multimodal + exercises vs non exercise treatment</td>
<td>Strong evidence of benefit</td>
<td>Pain, function</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GPE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GPE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GPE</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Studies included</td>
<td>Interventions</td>
<td>Evidence</td>
<td>Outcome</td>
<td>Remarks</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Schonstein 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No neck pain trials included</td>
</tr>
<tr>
<td>Back and neck</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peloso 2006</td>
<td>32</td>
<td>Methylprednisolone within 8 hours of accident (WAD) and IM lidocaine injections into myofascial triggerpoint for chronic MND</td>
<td>Moderate evidence of benefit</td>
<td>Pain</td>
<td>High quality positive trials not present.</td>
</tr>
<tr>
<td>Chronic neck</td>
<td></td>
<td>Epidural injections with methylprednisolone and lidocaine</td>
<td>Limited evidence of benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pain</td>
<td></td>
<td>Muscle relaxants and NSAID’s Botox-A intramusculair injections</td>
<td>Unclear benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trinh 2003</td>
<td>10</td>
<td>Acupuncture for chronic neck pain</td>
<td>Moderate evidence compared to sham post treatment and short term</td>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>Neck disorders</td>
<td></td>
<td>Limited evidence more effective than massage short term</td>
<td>Moderate evidence compared to waiting-list controls in short term</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Studies included</td>
<td>Interventions</td>
<td>Evidence</td>
<td>Outcome</td>
<td>Remarks</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Karjalainen 2003</td>
<td>2</td>
<td>Active multidisciplinary treatment vs traditional care</td>
<td>1</td>
<td>No evidence of benefit</td>
<td>Non randomised trial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multidisciplinary treatment with psychologist vs multidisciplinary treatment</td>
<td>1</td>
<td>Limited evidence of benefit</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>with psychologist as supervisor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kay 2005</td>
<td>31</td>
<td>Various types of strengthening, stretching and eye-fixation exercises</td>
<td></td>
<td>Limited evidence of benefit</td>
<td></td>
</tr>
<tr>
<td>Mechanical neck disorders</td>
<td></td>
<td>Active range of motion exercises, home exercise program</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eye fixation program</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stretching and strengthening</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multimodal approach, exercise combined with manipulations or mobilisations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pain, function, disability, patient</td>
<td>19 % van Tulder criteria to</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>satisfaction and GPE</td>
<td>35 % Jadad criteria were</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>high quality trials</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pain, functions, GPE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strong evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>evidence subacute and chronic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MND with or without headache</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>in short and long term</td>
</tr>
<tr>
<td>Study</td>
<td>Studies included</td>
<td>Interventions</td>
<td>Evidence</td>
<td>Outcome</td>
<td>Remarks</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Kroeling 2005</td>
<td>11</td>
<td>PEMF versus placebo</td>
<td>4</td>
<td>Limited evidence immediately post treatment chronic MND and acute WAD</td>
<td>Pain, improve function and disability</td>
</tr>
<tr>
<td>Mechanical neck disorders</td>
<td></td>
<td>Galvanic current, iontophoresis, TENS, PEMF long term effect</td>
<td>8</td>
<td>Unclear or conflicting evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diadynamic current, permanent magnets and electrical muscle stimulation</td>
<td>3</td>
<td>Limited evidence of no benefit</td>
<td></td>
</tr>
<tr>
<td>Haraldson 2006</td>
<td>19</td>
<td>Multimodal application or massage as stand alone treatment</td>
<td>6</td>
<td>Inconclusive evidence no recommendation can be made</td>
<td>Pain, physical functioning, patient satisfaction and cost of care</td>
</tr>
<tr>
<td>Kroeling 2005</td>
<td>11</td>
<td>PEMF versus placebo</td>
<td>4</td>
<td>Limited evidence immediately post treatment chronic MND and acute WAD</td>
<td>Pain, improve function and disability</td>
</tr>
<tr>
<td>Mechanical neck disorders</td>
<td></td>
<td>Galvanic current, iontophoresis, TENS, PEMF long term effect</td>
<td>8</td>
<td>Unclear or conflicting evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diadynamic current, permanent magnets and electrical muscle stimulation</td>
<td>3</td>
<td>Limited evidence of no benefit</td>
<td></td>
</tr>
</tbody>
</table>

GPE = Global Perceived Effect  
MND = Mechanical Neck Disorders  
WAD = Whiplash Associated Disorder
An update of conservative treatment and cost-effectiveness in neck pain

Economic evaluations of neck pain

Five economic evaluations of RCTs have been published on cost-effectiveness and cost-utility for conservative treatments of neck pain\textsuperscript{13;16;20;28;29;34}. Two of these studies\textsuperscript{13;28;29} did not exclusively include patients with neck pain nor report results separately for neck pain. Despite this, these studies are included in the summary below. The first Swedish economic evaluation compared chiropractic and physiotherapy for patients with low back- or neck pain visiting a general practitioner\textsuperscript{28;29}. In total, 323 patients aged 18 to 60 years who had no contraindications to manipulation and who had not been treated within the previous month were randomized to chiropractic (n=179) or physiotherapy (n=144). Treatment was carried out at the discretion of the therapist. Both direct and indirect costs were measured. There were no differences in outcome or direct or indirect costs between chiropractic and physiotherapy after 6 and 12 months. However, only 22\% of the patient population in this study had neck pain. A recent study conducted in the Netherlands compared the cost effectiveness of physiotherapy, spinal mobilisation, and usual care by a general practitioner(GP) for patients with neck pain. Patients were recruited by 42 general practitioners and randomly allocated to manual therapy (n=60, spinal mobilisation), physiotherapy (n=59, mainly exercise), or general practitioner care (n=64, counseling, education, and drugs). Both direct and indirect costs were prospectively measured using cost diaries covering a period of one year. The manual therapy group showed a faster improvement than the physiotherapy group and the general practitioner care group up to 26 weeks, but there were no differences in effectiveness after 52 weeks\textsuperscript{12}. The total costs of manual therapy (447 euro) were approximately one-third of the costs of physiotherapy (1297 euro) or the general practitioner care (1379 euro). The cost-effectiveness and cost-utility ratios showed that manual therapy was less costly and more effective than physiotherapy or general practitioner care\textsuperscript{5}. See figure 1 and 2.
Figure 1: Differences in results between manual therapy, physiotherapy and usual care by the general practitioner (Hoving et al 2002)

Figure 1 indicates the difference in overall improvement between the three intervention groups. The difference between manual therapy and physiotherapy or standard general practice was statistically significant. Results were similar for all other outcomes. Figure 2 shows a cost-effectiveness plane. The graph represents bootstrap replications of cost-effectiveness ratio for pain intensity comparing manual therapy with GP care. Most cost-effectiveness ratios are located in the bottom right quadrant, suggesting that manual therapy is more effective and less expensive than GP care.

Jensen et al\textsuperscript{13} conducted an economic evaluation of behavior oriented physiotherapy, cognitive behavioral therapy, a multidisciplinary rehabilitation program combining both, and usual care for patients with back and neck pain.
The study population consisted of blue-collar and service/care workers on sick leave, identified in a national health insurance database in Sweden. Approximately 40% of the study population had neck pain. Outcome variables were sick leave, early retirement and health-related quality of life. Both direct and indirect costs were included. The results showed that the multidisciplinary rehabilitation program was superior to the three other interventions, especially in women. However, data on health-related quality of life were not analyzed because of the low response rate and a formal cost-utility analysis was consequently not performed.

A German study assessed the costs and cost-effectiveness of additional acupuncture treatment in patients with chronic neck pain compared to patients receiving ‘usual care’. Both direct and indirect costs were included. Since health insurance databases were used, direct costs outside the health care system, for example over-the-counter medication, were not included. Primary outcome was health related
quality of life (SF-36). Follow-up was 3 months. A total of 1,753 patients were randomized to acupuncture and 1,698 to usual care. The costs of acupuncture treatment were significantly higher compared to usual care (926 Euro vs. 648 Euro; mean difference: 278 Euro [95% CI: 176 Euro to 379 Euro]). The incremental cost-effectiveness ratio (ICER) of acupuncture treatment was 12,469 Euros per QALY. The ‘usual care’ group included delayed acupuncture treatment after 3 months, which may not be an optimal usual care control group.

An economic evaluation conducted in the United Kingdom assessed the cost-effectiveness of a brief physiotherapy intervention versus usual physiotherapy management of neck pain\textsuperscript{20}. A total of 139 patients were allocated to the brief intervention, and 129 to the usual physiotherapy. Only direct costs were included and resource use data were prospectively collected for the follow-up period of one year. Quality-adjusted life years (QALYs) were estimated using EQ-5D data collected at baseline, 3 and 12 months from the start of the treatment. The results showed that the brief intervention was associated with lower costs (-68£; 95% CI -103£ to -35£). There were no differences in QALYs (-0.001; 95% CI, -0.030 to 0.028) compared with usual physiotherapy. The cost-utility ratio showed that the incremental costs of usual physiotherapy compare to the brief intervention were 68,000£ per QALY.

**Discussion**

In summary, the Cochrane reviews conclude that there are few high quality trials, that the effectiveness of many commonly used conservative treatments for neck pain is still unclear, that there are many small trials, and that effect estimates are also small. Manipulation and/or mobilization when used in combination with exercises seems the most promising. Overall, this seems a poor basis to establish clinical guidelines. Which technique or dosage was more beneficial was not possible to determine, neither whether certain subgroups benefit more from exercises than another subgroup.

In addition to an overview of the literature, good quality systematic reviews also appraise the methodological quality. However, clinical relevance of the trials is often ignored. In the neck pain reviews, there was hardly any focus on the content of the therapy used, failing to describe which techniques were used and if they were properly performed. Another problem related to the applicability of trial results, is the fact that many interventions evaluated in trials consist of a combination of different interventions or components. As a result, it is often impossible to assess which component of the therapy was
successful and why. Additionally, there is little agreement as to what manipulation, manual therapy, and mobilization encompass. We argue that future reports of trials and reviews should spend more attention to aspects of clinical relevance, and clearly describe the type, content and duration of the intervention. Despite the fact that the content of the interventions in the trials varied widely, the conclusion of this overview of reviews is that manual therapy, i.e. manipulation and/or mobilization, seems to be an effective therapy.

Additionally, this overview has shown that there is a conspicuous absence of high quality trials, especially for WAD. For this reason, guidelines for WAD in the Netherlands and the UK have been based upon expert opinion and on clinical experience of respected authorities in the field instead of on high quality evidence.

Finally, economic evaluations on patients with WAD or mechanical neck pain are rare. The economic evaluations that have been published showed that manual therapy and a brief physiotherapy intervention, might be more cost-effective than physiotherapy alone, and that acupuncture might be more cost-effective than usual care with delayed acupuncture treatment. The economic evaluations have been conducted in five different countries and results may not be directly generalizable to other countries, because of differences in health care and social systems. We argue that within the framework of EBM there should be more attention on economic evaluations because they give additional information on costs and the consequences of new or existing interventions\(^7\)\(^9\). Given budgetary limitations, it is not only important to know whether an intervention is more effective than another intervention, but also whether this is associated with lower costs.

Conclusions

The evidence on neck pain still is inconclusive and scarce, therefore, recommendations are usually based upon expert opinion rather than high quality studies. Randomized trials and economic evaluations that have sufficient sample sizes and meet current methodological standards are direly needed. The content of the interventions must be an integrated part of the description of these future trials, so they are more transparent, reproducible and their results generalizable to daily practice. This will also facilitate their role in clinical guidelines and EBM.
Reference list

28. Skargren EI, Carlsson PG, Oberg BE. One-year follow-up comparison of the cost and effectiveness of chiropractic and


Comparison of the effectiveness of a behavioural graded activity program and manual therapy in patients with sub-acute neck pain: design of a randomised clinical trial

Published as;

Chapter 3

Abstract

The objective is to present the design of a randomised clinical trial (RCT) on the effectiveness and cost-effectiveness of a behavioural graded activity programme compared with manual therapy in patients with sub-acute neck pain. Sub-acute is defined as pain existing for 4 to 12 weeks. The behavioural graded activity programme is a time-contingent increase in activities from baseline towards pre-determined goals. Manual therapy consists mainly of specific spinal mobilisation techniques and exercises. The primary outcomes are global perceived effect and functional status. Secondary outcomes are kinesiophobia, distress, coping, depression and somatisation. The intensity and persistence of the pain and its interference with activities are also assessed. Direct and indirect cost are measured by means of cost diaries. Measurements take place at baseline and 6 and 12 weeks after randomisation. To assess the long term effect, measurements will also take place after 6 and 12 months. Finally some challenges are discussed concerning the use of a behavioural graded activity programme, manual therapy and outcomes.
Introduction

Neck pain is a common musculoskeletal disorder. The point prevalence for neck pain in the general population of the Netherlands varies between 9% and 22%, and approximately one-third of all adults will experience neck pain during the course of 1 year. Some 5-10% of the neck complaints will develop into a chronic pain disorder. Once non-specific neck pain becomes chronic, defined as pain existing for more than 12 weeks, 44% of the patients consult their general practitioner annually. The main feature of neck pain is pain in the cervical region, often accompanied by restriction of the range of motion and functional limitations. The pain may originate from many structures in the cervical region, especially the spine and soft tissues, but there is no conclusive evidence regarding specific pathology in the majority of cases of acute or chronic neck pain. Consequently most cases are labelled as non-specific neck pain or neck pain of unknown origin. Risk factors for the occurrence of neck pain are physical load factors such as vibration, flexion of the neck, sitting posture and heavy lifting, but psychological and social factors are also reported to aggravate and perpetuate neck pain. Hence, neck pain is a biopsychosocial problem, in line with the definition of pain formulated by the International Association for the Study of Pain (IASP 1986).

High pain intensity and a previous history of neck pain is strongly and consistently associated with an unfavourable prognosis. Although neck pain is often self-limiting within a few weeks of onset, 40% of the patients contact their general practitioner (GP); 30% are referred for further diagnosis to a medical specialist and 32% for physiotherapy, manual therapy or some other type of conservative therapy. The evidence regarding the effectiveness of these conservative therapies for neck pain is still inconclusive. A review performed by Aker et al showed no benefit from stretching, laser therapy, traction, exercise or neck school for acute neck pain. Gross et al concluded that manual therapy is effective for neck pain in the short term, if used in combination with types of other treatment (e.g. exercises). The updated Cochrane review of Gross et al concluded that there was a strong evidence for manipulation and/or mobilization when used in combination with exercises. Manipulation and/or mobilisation alone were not beneficial. For chronic neck pain, randomised controlled trials (RCTs) have reported beneficial effects in favour of physical therapy, acupuncture and manual therapy. Bronfort concluded that their was moderate evidence that spinal manipulative therapy and/or
Chapter 3

mobilisation was superior to general practice medical care and physical therapy in the short term for improving physical function in patients with chronic neck pain. For patients with acute neck pain the evidence was inconclusive. Hoving et al\textsuperscript{26} concluded, after an extensive review of reviews, that there is no conclusive evidence for or against any of these treatments. However, recently published results of an RCT carried out by Hoving et al\textsuperscript{27} in patients with sub-acute or chronic neck pain showed a significant difference in effectiveness in favour of manual therapy, compared to both physiotherapy or usual care from the GP, both for short and long-term follow-up.

One of the shortcomings of a review seems to be the focus on the methodological quality of the trials. The quality of the trials is in most cases poor and there is hardly any focus on the content of the therapy used. If the definition of manipulation and/or mobilisation is common, trials are included in the review without knowing if the used techniques are properly used, if there is a treatment protocol and if the techniques are generally used in daily practice.

Despite that manual therapy seems to be an effective therapy. Among some patients neck pain still becomes chronic. One possible explanation could be the role of psychological and social factors in the awareness of pain. During recent decades there has been an increasing interest in the psychological and social aspects of acute and chronic pain. In addition, psychological and social factors are believed to play a role in the transition from acute to chronic pain and disability\textsuperscript{17;39}. Consequently for patients with sub acute and chronic pain the emphasis is increasingly focussed on behavioural treatment, based on operant, cognitive or respondent techniques\textsuperscript{44;55;58;61}. Behavioural treatment focuses on reducing disability through the modification of environmental contingencies and cognitive processes.

Also a transition of a similar trend can be observed for sub-acute and chronic neck pain. Identification of the underlying specific pathology is no longer the primary focus. For this, several reasons are mentioned: a) medical examinations fails to find specific underlying pathology in the majority of neck pain cases\textsuperscript{8}, b) the degree of physical disability can be due to inactivity rather than a result of the physical condition\textsuperscript{33}, c) the pain can depend on cognitive processes\textsuperscript{11} and negative thoughts\textsuperscript{14}, and d) the patient’s condition can depend on the degree of kinesiophobia\textsuperscript{35}. This model suggest possible pathways by which neck pain patients, similarly to low back pain patients or patients with other pain conditions, become enmeshed in a downward spiral of increasing avoidance, disability and pain. Specially in patients who interpret pain as threatening (pain catastrophizing) and exhibit kinesiophobia or fear of
Design of a RCT comparing effectiveness of graded activity and MT movement. In literature this model has been a topic for research specially, in low back pain patients but not in neck pain patients so far. Although there are some promising results regarding the effect of cognitive behavioural therapy (CBT) on back pain, arthritis pain, cancer pain and mixed chronic pain, the effectiveness of CBT for neck pain in a primary care setting is still unknown. Therefore, we hypothesize that the above mentioned factors are also involved in neck pain patients and we suggest that CBT is also a useful therapy for patients with sub-acute neck pain.

A CBT programme is based on the bio-psychosocial model, which means that not only the nociceptive structures are held responsible for the pain awareness of the patient. Pain can also been seen an emotion according to the IASP (International Association for the Study of Pain) definition of pain. This can lead to a response in one of the following three response systems that characterise emotional experiences: i) the psycho-physiological system such as feelings, increase muscle tension etc.; ii) the cognitive system, such as thoughts, catastrophizing, fear, etc.; and iii) the motor system such as pain behaviour, disuse syndrome, etc. Physical therapists are not trained to treat cognitive processes, so a full CBT program is not realistic. Pain behaviour however can be treated by PT’s using a graded activity (BGA) programme as incorporated in other trials. The focus is on the motor system and the PT can use operant principles and can act as a coach. The evidence of this BGA programme is still questionable, however it is widely practiced in low back pain patients. Some studies are not positive, others are more promising in improving the level of physical activity at work compared to usual care. However for neck pain patients the affect of a BGA programme is still unknown. In our opinion it is a challenge to assess the effectiveness of this programme in patients with sub-acute neck pain.

In summary, manual therapy, a typical hands-on therapy is an effective therapy for neck pain. It is hypothesized that psychological and social aspects play an important role in the transition from sub-acute to chronic pain. BGA, a typical hands-off therapy, can influence pain behaviour and pain intensity by focussing on those aspects, and shows promising results in other pain conditions. In order to assess the effects of BGA for neck pain we designed an RCT assessing the following hypothesis:

A BGA programme is more effective than manual therapy in patients with sub-acute neck pain.

Secondary we will assess whether the severity of complaints influences.
Chapter 3

The study protocol was approved by the Medical Ethics Committee of the VU University Medical Centre in Amsterdam.

Methods

Selection of patients

The participants in the study are patients with sub-acute non-specific neck pain, defined as pain in the cervical region existing for at least 4 weeks, but no longer than 12 weeks. The neck pain may radiate to the shoulder region or the upper extremities, or be accompanied by headache, but the main complaint must concern the neck. The inclusion criteria are: non-specific neck pain, age between 18 and 70 years, and a new episode of pain (defined as no neck pain in the previous 4 months). The patients must not have had any therapy for neck complaints in the previous 4 months. The exclusion criteria are specific neck pain, for example due to rheumatoid arthritis, disc herniation, neurological diseases or malignancy. Patients with whiplash-associated disorders are included unless they have an unsettled insurance claim running during the intake period. During the first GP consultation these criteria are assessed and the patient is informed about the study. Eligible patients who are interested in participation are referred to the research assistant, who informs them further about the consequences of participation and re-checks the inclusion criteria. Patient who are eligible and agree to participate are asked to sign the informed consent form and the baseline measurement is performed.

Randomisation procedure

After the baseline measurement the patients are randomly assigned either to the manual therapy treatment (MT) or to the BGA programme. The treatment sessions take place in the private practices of the participating therapists. A colleague from the research department (RO) who is not involved in recruitment, treatment or data-collection, generated a random list based on a computer-generated sequence. The randomisation was pre-stratified for severity of the complaints and age of the patient. Four strata are constructed with a cut-off point for age of 40 years and a cut-off point for severity of the main complaints of 7 on a 0-10 numerical rating scale. The treatment allocation is concealed, as numbered opaque sealed envelopes based on the computer generated list are used, and the research assistant who deals with the inclusion of the patients, is unaware of the content of the envelopes.
Blinding

The patients are aware of the treatment they receive, so it is not possible to blind them but the research assistant who is responsible for the baseline and the follow-up measurements will be blinded for the treatment allocation. Prior to the measurements, the patients are asked by the research assistant not to mention the treatment to which they were allocated. To evaluate the blinding procedure, at the end of the follow-up period the research assistant will record which treatment she thinks the patients received.

Drop-outs

All efforts will be made to avoid drop-outs, such as extra telephone calls and/or mails and if necessary a visit at the patient’s home address.

Sample size

Although arbitrary this is based on the expectation that in the manual therapy group 70% of the patients will recover. To detect a difference of 20% between the two treatment groups, which is considered as clinically important, 84 patients are required for each treatment group. This calculation is based on the dichotomised primary outcome measure “perceived recovery”, defined as the percentage of patients who are reported to have recovered. The sample size calculation concerns an α of 0.05 and a power (1-β) of 90%. To compensate for drop-outs during follow up, we planned to include 90 patients per treatment group. To obtain the required sample size, patients will be recruited by 70 GP’s.
Figure 1 The study design, patient flow.

BGA: Behavioural Graded Activity programme
MT: manual therapy
Description of the interventions

Manual therapy

In the Netherlands, manual therapy is a specialisation of physiotherapy. The manual therapists in this study followed 3-year post-graduate courses in manipulation and/or specific mobilisation techniques to become certified and registered by the Royal Dutch Society for Physical Therapy (KNGF) as a manual therapist. The aim of manual therapy is to recognize and interpret tissue and organ-specific dysfunctions on a local and segmental level. During the physical examination, the musculoskeletal system is examined, while accepting asymmetrical morphology and function and respecting the related individual preference of function. A biomechanical assessment is made to obtain detailed information about the relevant joints, muscles, and surrounding soft tissue. The assessment of the cervical spine includes three-dimensional tests within or at the limit of the range of motion of the joints. The aim of the treatment is to restore restricted movement, stimulating natural recovery and adaptive processes in relation to the functionality of movement. Furthermore, the treatment also aims to reduce pain, to increase the patient’s level of activities and participation, and to prevent recurrences.

The treatment consists of manipulation and specific mobilisation techniques. A manipulation is a passive movement of a joint beyond its active and passive limit of motion, but within the limit of its anatomical integrity. It is usually a localised thrust which is a quick movement of small amplitude led by the therapist. The aim of the manipulation is to regain motion, to restore function and to reduce pain.

A mobilisation utilises skilled low-grade passive movement with large amplitude. Passive mobilisation can be repetitive or not, varying in amplitude. The aim of mobilisation is to restore movement and to relieve pain. The specific technique that is chosen depends on the therapist, and is not yet a topic for research. Similar to the Hoving trial, manual therapy did not include high velocity thrust techniques in the cervical region. This technique was excluded for ethical reasons because of reported complications of spinal manipulations, and especially vertebrobasilar complications. Despite this exclusion, the overall effect of the manual therapy intervention was promising in favour of manual therapy. Dutch manual therapists uses knowledge, methods, and techniques considered unique to manual therapy. In daily
clinical practice, physical therapy and manual therapy are often less distinct, because the same person, i.e., the physical therapist with a specialization in manual therapy, provides both. So it is standard practice to use additional exercises and give advice as well in a manual therapy treatment. These are patient tailored and the aim of the exercises are mobilisation, stabilisation and coordination. This is illustrated by Table 1 which shows the registered content of the manual therapy in the Hoving trial\textsuperscript{27}.

The content of manual therapy in this trial will be the same as in the Hoving trial.

Table 1: The content of manual therapy treatment, assessed by manipulative therapist on a registration form\textsuperscript{27}.

<table>
<thead>
<tr>
<th>Manual therapy in neck pain patients (n=60)</th>
<th>Median (IQR)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sessions</td>
<td>6 (5, 6)</td>
</tr>
<tr>
<td>Physical examination</td>
<td>6 (5, 6)</td>
</tr>
<tr>
<td>Muscle techniques</td>
<td>5 (3, 6)</td>
</tr>
<tr>
<td>Specific articular mobilisation techniques</td>
<td>6 (4, 6)</td>
</tr>
<tr>
<td>Frequently used:</td>
<td></td>
</tr>
<tr>
<td>- Type: traction /translation,</td>
<td></td>
</tr>
<tr>
<td>2/3 dimensional specific techniques</td>
<td></td>
</tr>
<tr>
<td>- Location: 2\textsuperscript{nd} – 3\textsuperscript{rd} cervical segment, cervico-thoracic junction-thoracic spine and costo-vertebrae articulations (1st rib)</td>
<td></td>
</tr>
<tr>
<td>Co-ordination and stabilisation techniques</td>
<td>3 (0, 4)</td>
</tr>
<tr>
<td>Instruction and exercises</td>
<td>4 (3, 6)</td>
</tr>
</tbody>
</table>

* IQR (Inter Quartile Range of treatment sessions)

In summary; in the current RCT the MT intervention is similar to the intervention described in the Hoving trial, which consisted of manual therapy techniques, exercises and advice, we will refer to this therapy as MT. The therapists are also asked to fill in a registration form after each session. The therapists are allowed to provide a maximum of six treatment sessions within six weeks. The duration of a single treatment session is 30-45 minutes.
Behavioural graded activity programme (BGA)

To emphasize the behavioural component, compared with physical training, the term behavioural graded activity was introduced. In general, the focus of the treatment is on function and not on the underlying pathology or biological aspects of pain. Physiotherapist are already trained to treat functional recovery in patients with all kinds of limitations, and they are properly equipped to conduct the intervention in the current trial after a special 2-day training. The PT’s participating in this trial all have had additional courses in biopsychosocial approach of pain problems. All PT’s had more than 10 years of clinical experience. To solve all existing problems and to monitor the treatments, regular days of reflection are organised, furthermore registration forms are used to get insight in the treatments. The therapists attitude is checked using a health care providers questionnaire which was adapted for neck pain.

The BGA programme, as applied in the present study, is based on time-contingent management, as described in more detail by Fordyce and applied by Lindstrom and Ostelo. The emphasis of the treatment in this trial is the operant strategy. Core elements of this programme are: 1) decrease in the pain behaviour and increase in "well" or “healthy” behaviour, 2) improving function and not the reduction of pain, 3) the patient is responsible for the treatment and has an active role, and 4) the therapist acts as a coach.

The BGA treatment can be divided into three phases which will be discussed separately.

1) Initial phase

The initial phase first concerns a reconceptualisation of the patient’s pain model. Central in this is the understanding that pain is not solely the result of underlying tissue damage, but is also influenced by the patient’s expectations, beliefs, and fear, as well as activity levels and home and work environment. The patient is then taught that it is safe to move the cervical spine or other parts of the body. Subsequently, the three main complaints are formulated at baseline. A main complaint is defined as an activity that is very important to the patient, implying that improvement of these activities is highly desirable. During the initial phase the patient is asked to perform these activities until the pain becomes too dominant, in other words pain-contingent. The level,
Chapter 3

duration or frequency of activities, is registered on a performance chart. A baseline level is constructed, based on these performance charts, thereby determining the average level of each specific activity. From baseline level the patient has to set his/her own individual treatment goals. For example, the patient wants to be able to read his documents for 12 minutes during work (see example). Once the goal is determined, and knowing the baseline level, quotas are set in order to achieve this predetermined treatment goal within a predetermined time-span.(time-contingent)

2) Treatment phase

Once the treatment phase starts, activities and exercises are no longer performed on a pain-contingent basis, but follow the predetermined quotas. Therefore the key element of the treatment phase is time-contingency, meaning that despite the pain or discomfort the quotas will be adhered to. Initial quotas are set in such a way that they are slightly below the baseline level, to ensure that the first treatment session will be a successful one. During treatment, the therapist stimulates and encourages the patient and gives positive reinforcement if the quotas are achieved.

3) Generalisation phase

The aim of this phase is to encourage the patients to proceed with their healthy behaviour during activities of daily living. It is not sufficient to train in a treatment setting only, but one should also generalise the goals into working or home situations. In this phase the frequency of treatment sessions will be diminished and self-efficacy will be strongly encouraged.

The BGA treatment within the trial period will consist of a maximum of 18 sessions of approximately 30 minutes.
Patient A suffers from neck pain and is not able to read his documents during work. The patient is asked to score this activity during five days of work. The question is; what time is he able to read his document till the pain in his neck starts to come. The first day this will be after 3 minutes, the second day after 6 minutes, the third day after 3 minutes the fourth day after 8 minutes and the fifth day after 4 minutes. This is the pain contingency phase. To start the treatment phase a mean is constructed, here 5 minutes. The treatment starts with a exercise of 2 minutes of reading to assure a positive and not a painful start. In consensus with the wish of the patient the quotas in this example are 2 minutes every 2 days with a therapy goal of 12 minutes in 16 days. The treatment phase is characterized by its time-contingency. The pain is not an issue in completing this predetermined treatment goal.

**Figure. 2** Example of constructing a baseline measurement of a main complaint.

**Measurements**

The demographic variables as well as primary and secondary outcomes are measured at baseline. Table 2 gives an overview of the data-collection.
Chapter 3

Demographic variables
Demographic variables, such as age and gender, will be registered. Furthermore, disease characteristics such as history of the neck pain, possible cause of the complaint, duration of the complaint, irradiation to shoulder or extremity, accompanying headache, shoulder or back pain, will be assessed.

Primary outcome measurements
Global Perceived Effect\(^6\)\(^{15}\) is measured by self-assessment on a 7-point scale, 1= completely recovered, 2= much improved, 3= little improvement, 4= no change at all, 5= slightly worse, 6= much worse and 7= worse than ever. The neck-specific functional status is measured according to the Neck Disability Index\(^59\). The Dutch translation was found to be a sufficient validly instrument\(^24\).

Secondary outcome measurements
These measurements will evaluate all domains of the psychological and social aspects of pain defined in the introduction. Fear of movement is measured according to the Tampa Scale for Kinesiophobia (TSK)\(^36\). The Dutch translation of the TSK has a fair and consistent internal validity\(^19\). Pain catastrophising, pain coping and pain control (external and internal control) is measured by means of the Pain Coping and Cognition List (PCCL)\(^51\). The 4 Dimensional Symptom Questionnaire (4DSQ) measures factors such as distress, depression, fear and somatisation as intermediate factors\(^52;53\). Within the Numeric Rating Scale (NRS) for pain, the patients score their average and maximum pain in the past week and current pain on an 11-point rating scale. The NRS is a valid and responsive scale\(^47\).

The Graded Chronic Pain Scale (GCPS) is designed to assess the intensity, interference with activities and persistence of pain. In the current trial it is used to assess neck pain. All items are either scored on a NRS scale or expressed in days\(^52;53\). The Patient Specific Questionnaire\(^7;6;32\) is used to score the three most important disabilities on an 11-point numerical rating scale (0 no disability -10 not able to perform this activity).

Health status is evaluated with the Short Form 36 (SF-36). The Dutch translation showed satisfactory validity and reproducibility\(^1\). Quality of life is measured according to the Euroqol-5D\(^15;54\). Furthermore, the patients will record any costs due to their neck pain, visits to the therapists, absenteeism from work and use of medication, in a cost diary\(^18\).
Table 2: Overview of data collection

<table>
<thead>
<tr>
<th>Instrument</th>
<th>baseline</th>
<th>6wks</th>
<th>12wks</th>
<th>26wks</th>
<th>52wks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion/exclusion</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Recovery</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Neck Pain Disability Index (NDI)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Secondary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tampa Scale of Kinesiofobia (TSK)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pain Coping and Cognition List (PCCL)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4 Dimensions of Psychological Symptomatology Questionnaire (4DSQ)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Numerical Rating Scale for pain (NRS)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patient Specific Questionnaire (PSQ)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Graded Chronic Pain Scale (GCPS)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Sort Form 36 (SF-36)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>EuroQol</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cost-diary</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Analysis

The baseline scores of the patient’s demographic (e.g. age, gender, duration of complaints, history of complaints and trauma), primary and secondary outcomes will be used to compare the two intervention groups. Differences between baseline and follow-up measurements will be calculated, and compared between the two intervention groups. If necessary, adjustments for baseline variables will be made, using analysis of covariance. Considering the longitudinal context of the data and possible confounding on the level of the therapist a generalized linear mixed model will be used. The statistical analyses will be performed on the basis of the intention-to-treat principle, i.e. patients
will be analysed in the treatment group to which they were randomly allocated. Missing data will be replaced by a linear interpolation method for missing measurements, and by a ‘last measurement carried forward’ method for drop-outs. Worse-case scenarios and best case-scenarios for patients with and without missing values for the end-point variables will be compared for the total study population and per treatment group.

Discussion

Publishing the design of a study before publishing the results is important for several reasons. Firstly, it yields an opportunity to reflect critically on the design, independently of the study outcomes. Secondly, if the design is published any deviations from the original design can be identified when the study results are published. Thirdly, it may counteract possible publication bias, because authors of future systematic reviews can identify the study even if its results are never published.

Manual therapy seems to be an effective treatment for patients with neck pain (recent reviews of literature). But the exact content of manual therapy is not always clear. In the Netherlands manual therapy consists of specific manual techniques, exercises and advice and is frequently used to treat patients with neck pain. This approach seems to be an effective treatment for these patients. However, in a majority of patients with neck pain the complaints are persistent or have recurrences. Cognitive and behavioural factors seem to play an important role. BGA focuses on these factors, but the evidence of the effectiveness of BGA is lacking. Therefore a randomised controlled design was designed to compare the effectiveness of BGA versus manual therapy.

Challenges of this design are;

The treatment BGA. The behavioural graded activity program used in this trial is an operant therapy based on the principles of the bio psychosocial model. Although physical therapist are skilled to treat patients with neck pain it is not self evident that they are able to provide a BGA program and change their attitude from a pain contingent approach to time contingent approach. To ensure that these principles are adequately used a two day training program is provided, supervised by an experienced behavioural therapist and a psychologist. This program consist of a theoretical part
Design of an RCT comparing effectiveness of graded activity and MT

in which all the principles of a BGA program are discussed, and a practical part. Although it might be desirable to train PT’s more extensively, we choose to train PT’s according to the training courses that are normally provided in this approach. The advantage of this strategy is that if this trial provides evidence in favour of the BGA, the results can easily be implemented. The use of the BGA program by the therapists is evaluated during the trial using a registration form.

Manual therapy.
There is an ongoing discussion concerning the content of manual therapy used in trials. The reaction of many readers of BMJ on the article of the Hoving trial confirms this. In the Dutch situation manual therapy is a combination of manipulative therapy, specific mobilisation techniques, exercises and advice. In this trial the different components of manual therapy will be described in detail, which will benefit the interpretation of the results.

Outcome measures
Patient satisfaction can be measured using different scales or questionnaires. We choose Global Perceived Effect (GPE) as a primary outcome measure even though there are some concerns about the reliability and validity of global rating scales. Global ratings typically are correlated with the patient’s present status and are not an unbiased measure of change. However most authors regard global rating scales as clinically relevant and valid and responsive to measure patient’s perceived recovery. From the patient’s point of view this subjective scale is perhaps the most sensible method of assessment.

This study is designed as a RCT. The first challenge is to investigate whether BGA is more effective than manual therapy, with focus on a comparison between a mainly hands-on approach, based on the biomedical model, and a hands-off approach. The second challenge lies in the fact that the study population consists of patients with sub-acute neck pain. The behavioural approach has mainly been tested in chronic pain patients, in whom it is expected that psychological and social factors become more dominant over time. The turning point in pain behaviour, from more nociceptive dominance to more psychological and social dominance, is still unknown, although is it hypothesised that approximately 7-8 weeks after the onset the behavioural factors become dominant. So the question remains whether manual therapy or a
behavioural graded activity program can prevent sub acute pain patients from chronicity. The results of this RCT may be of value for the clinician in choosing the right therapy strategy for each individual patient. Furthermore, this RCT can be used to update systematic reviews, and may contribute to the development of evidence based clinical guidelines in this field.
Reference List


27. Hoving JL, Koes BW, de Vet HC et al. Manual therapy, physical therapy, or continued care by a general practitioner for patients
Design of an RCT comparing effectiveness of graded activity and MT


49. Staal B. Low Back pain, graded Activity and Return to Work. 2003. VU University Medical Centre Amsterdam.


Is a behavioural graded activity programme more effective than manual therapy in patients with sub-acute neck pain? Results of a pragmatic randomised clinical trial

Submitted as:
Jan Pool, Raymond Ostelo, Dirk Knol, Johan Vlaeyen, Lex Bouter, Riekie de Vet. Is a behavioural graded activity programme more effective than manual therapy in patients with sub-acute neck pain? Results of a randomised clinical trial
Abstract

Objective. Neck pain is a common complaint, for which many conservative therapies are available in primary care. Our objective was to compare the effectiveness of a behavioural graded activity programme (BGA) with manual therapy (MT) in patients with sub-acute (4-12 weeks) non-specific neck pain.

Methods. A randomised clinical trial was conducted, involving 146 patients with sub-acute non-specific neck pain, recruited by 35 general practitioners. The BGA programme can be described as a time-contingent increase in activities from baseline towards pre-determined goals. Manual therapy consists mainly of specific spinal mobilisation techniques and exercises. Primary outcomes were global perceived effect (GPE), the Numerical Rating Scale (NRS) for pain intensity and the Neck Disability Index (NDI). Secondary outcomes were the Tampa Scale for Kinesiophobia (TSK), the 4 Dimensional Symptom Questionnaire (4DSQ) measuring distress, depression, fear and somatisation, and the Pain Coping and Cognition List (PCCL) measuring catastrophizing, coping and internal and external pain control. Measurements were carried out at baseline and 6, 13, 26 and 52 weeks after randomisation. Data are analysed according to the intention-to-treat principle, using multilevel analysis.

Results. The success rates at 52 weeks, based on the GPE were 89.4% for the BGA programme and 86.5% for MT. This marginal difference was not statistically significant. For pain and disability, a marginal difference was found in favour of the BGA programme; mean difference for pain = 0.99 (95% CI: 0.15-1.83) and mean difference for NDI = 2.42 (95% CI: 0.52-4.32). All other differences between the interventions in the primary and secondary outcomes were not clinically relevant or statistically significant.

Conclusions. Based on this trial it can be concluded that there are only marginal, but not clinically relevant, differences between a BGA programme and MT.
Introduction

Neck pain is a common complaint. The point prevalence of neck pain in the general population of the Netherlands varies between 9% and 22%;6,26, and approximately one-third of all adults will experience neck pain during the course of 1 year. Some 5-10% of these patients will develop a chronic pain disorder. Once the neck pain becomes chronic (> 12 weeks) 44% of the patients will consult their general practitioner (GP) in the following 12 months. There is no conclusive evidence regarding specific pathology in the majority of cases of acute or chronic neck pain so, consequently, most cases are labelled as non-specific neck pain or neck pain of unknown origin. Psychological and social factors may aggravate and perpetuate neck pain. The therapeutic modalities that are most frequently used for the treatment of neck pain are exercises, manipulative therapies, mobilisation, massage, physical modalities and multidisciplinary biopsychosocial rehabilitation. Although the evidence regarding the effectiveness of most conservative types of therapy for neck pain is still inconclusive, the updated Cochrane review carried out by Gross et al concluded that there was strong evidence for manipulation and/or mobilisation, if combined with exercises. Furthermore, the results of a recent randomised clinical trial (RCT) carried out by Hoving et al in patients with sub-acute and chronic neck pain showed a significant difference in effectiveness and cost-effectiveness in favour of MT, compared to physical therapy or usual care from the GP, both in the short and the long-term follow-up.

Psychological and social factors are believed to play a role in the transition from acute to chronic pain and disability. Consequently, the emphasis in treating patients with sub-acute and chronic pain is increasingly on behavioural treatment, based on operant, cognitive or respondent techniques. Behavioural treatment focuses on reducing disability through the modification of environmental contingencies and cognitive processes. Especially for patients who interpret pain as threatening (pain catastrophizing) and are afraid that movement might be harmful (kinesiophobia), behavioural treatment seems to be effective. This led to the development of a behavioural graded activity (BGA) programme. The evidence of the effectiveness of a BGA programme, which is still emerging, is mainly from studies on low back pain. However, the effectiveness of a BGA programme for neck pain is still unknown. We hypothesise that psychological factors such as those described above, will influence neck pain, and we suggest
that a BGA programme will be at least as effective as MT for patients with sub-acute neck pain.

**Methods**

The methods have been described elsewhere in detail.

**Selection of patients**

To detect a difference of 20% between the two treatment groups, we aimed to include 90 patients with sub-acute non-specific neck pain per treatment group. Sub-acute neck pain was defined as pain in the cervical region, existing for at least 4 weeks, but no longer than 12 weeks. The neck pain could radiate to the shoulder region or the upper extremities and/or be accompanied by headaches, but the main complaint must concern the neck. The other inclusion criteria were: age between 18 and 70 years, and a new episode of non-specific neck pain, defined as no neck pain in the previous 4 months. The exclusion criterion was specific neck pain, for example due to rheumatoid arthritis, disc herniation, neurological diseases, or malignancy. Patients with whiplash-associated disorders were included, unless they had an unsettled insurance claim during the intake period. During the first GP consultation these criteria were assessed and the patient was informed about the study. Eligible patients who were interested in participation were referred to the research assistant, who informed them about the consequences of participation and re-checked the inclusion criteria. Patients who were eligible and agreed to participate were asked to sign the informed consent form before the baseline measurement was performed. The baseline measurement consisted of collecting data on demographic variables, potential prognostic factors and outcome measurements.

The study protocol was approved by the Medical Ethics Committee of the VU University Medical Center in Amsterdam.

**Randomisation procedure**

After the baseline measurement the patients were randomly assigned either to the BGA programme or to MT. The treatment allocation was concealed, through the use of numbered, opaque and sealed envelopes, based on a computer-generated list, prepared by an independent person before the start of the inclusion period.
Blinding
The research assistant who was responsible for all measurements, and visited the patient at home, was blinded for the treatment allocation. The patients were aware of the treatment they received.

Behavioural graded activity programme
The physical therapists (PTs) who provided the BGA programme in this trial had followed additional courses in the biopsychosocial approach of pain problems, to make sure they had the necessary skills. Furthermore, all the PTs had more than 10 years of clinical experience. For this study they attended an additional 2-day training course, which consisted of a theoretical part, in which the principles of a BGA programme were discussed, and a practical part, in which skills were trained using case reports. The course supervised by an experienced behavioural therapist (AK) and a psychologist (JWSV). The BGA programme was based on time-contingent management, as described in more detail by Fordyce\textsuperscript{10,11} and applied by Lindstrom and Ostelo\textsuperscript{21,25}. The emphasis of the treatment is on operant conditioning. Core elements of a BGA programme are: 1) decrease in pain behaviour and increase in “healthy” behaviour, 2) improvement of function and no focus on pain reduction, 3) the patient is responsible for the treatment and has an active role, and 4) the therapist acts as a coach. The therapy is based on a typical “hands-off” approach.
From the baseline level, and working towards pre-set goals, the patients were trained to follow a gradually increasing exercise programme, which consisted of a maximum of 18 sessions of approximately 30 min.

Manual therapy treatment
The MT was provided by manual therapists. In the Netherlands after completing their training, PTs follow a 3-year post-graduate course in manipulation and specific mobilisation techniques to become certified and registered as an MT by the Royal Dutch Society for Physical Therapy (KNGF). The aims of the MT treatment are: 1) to restore restricted movement, stimulating natural recovery and adaptive processes in relation to the functionality of movement\textsuperscript{27}, 2) to reduce pain, 3) to increase the patient’s level of activities and participation, and 4) to prevent recurrences. The treatment consists of manipulation and specific mobilisation techniques. It is also standard practice to give additional exercises and advise during MT treatment. The therapist were allowed to provide a maximum of 6 treatment sessions, each with a duration of 30-45 minutes, within 6 weeks. The MT was similar to the MT provided in the Hoving trial\textsuperscript{16}. 
A detailed description of both interventions has been published elsewhere.27

Outcome measures

The baseline measurement consisted of data on patient characteristics, demographic variables, potential prognostic factors and outcome measurements (Table 1). Furthermore, the patients completed questionnaires at 6, 13, 26 and 52 weeks after randomisation. The Primary outcome measurements were: 1) Global Perceived Effect (GPE)4,9 measured on a 7-point scale ranging from ‘completely recovered’ to ‘worse than ever’ (recovery was a priori defined as ‘completely recovered’ or ‘much improved’, as reported by the patient); 2) the severity of current neck pain, scored on an 11-point Numerical Rating Scale (NRS) ranging from 0 ‘no pain’ to 10 ‘very severe pain’; and 3) the neck-specific functional status measured with the 10-item Neck Disability Index.35

The secondary outcomes were: 1) fear of movement, assessed with the Tampa Scale for Kinesiophobia (TSK);19 2) pain catastrophising, pain coping and pain control (external and internal control) based on the Pain Coping and Cognition List (PCCL)30; 3) distress, depression, fear and somatisation, assessed with the 4 Dimensional Symptom Questionnaire (4DSQ);31 4) the intensity of pain, interference with activities, and persistence of pain, assessed according to the Graded Chronic Pain Scale (GCPS);37,38; and 5) general health status, evaluated with the Short Form 36 (SF-36).1

Statistical analyses

The statistical analyses were performed according to the intention-to-treat principle. To determine the effectiveness of the interventions over the follow-up period, multilevel analyses were performed. In the analyses, patients clustered under therapists and repeated measurement clustered within a patient, were taken into account. Therefore, we used linear multilevel analyses for continuous variables, such as the NDI and the NRS, and logistic multilevel analyses for dichotomous variables, such as the GPE. We included the following levels: repeated measures (time), patient and therapist. The analyses took baseline scores into account and focused on the interaction between repeated measurement and intervention. The resulting regression coefficients (continuous variables) or odds ratios (dichotomous variables) can be interpreted as the difference in patient
outcomes between the two groups at a certain follow-up moment. Per protocol analyses were also performed, excluding all patients with deviations from the protocol. For the BGA programme a deviation of the protocol was defined as: not conducted the baseline level in the initial phase of the treatment, or suspension of therapy, applying ‘hands-on’ therapy, or using a pain-contingent approach. For the MT a deviation of the protocol was defined as not using manipulation or specific mobilisation techniques.

To calculate the sum scores, occasional missing items within a questionnaire (< 5%) were imputed, using a SPSS syntax for missing value imputation. When applying multilevel analyses to longitudinal data, no imputation strategy for missing questionnaires is necessary. It has been demonstrated that multilevel analyses are very flexible in handling these missing questionnaires. The analyses were performed in SPSS 12.0 and MLwiN 2.02.

Results

Participants
From January 2003 until January 2005 a total of 163 patients were referred to the research assistant. 146 patients met all inclusion criteria and signed the informed consent form. Figure 1 shows the flow chart of the study. The baseline characteristics of the patients in both groups were nearly similar (Table 1).
Chapter 4

**Figure 1.** Flow chart

Consultation with General Practitioner

Assessment of eligibility

Check inclusion, informed consent and baseline assessment at research centre (n=163)

Excluded (n=17);
- Pain for more than 12 weeks n=10
- Claim due to accident n=2
- Treatment in previous 6 months n=1
- Age < 18 n=1
- No complaints anymore n=2
- Dizziness as main complaint n=1

Randomisation (n=146)

- BGA treatment (n=71)
- MT treatment (n=75)

Follow-up: M* A#
- 6 weeks (n=68) (n=71)
- 13 weeks (n=68) (n=71)
- 26 weeks (n=60) (n=71)
- 52 weeks (n=66) (n=71)

Follow-up: M* A#
- 6 weeks (n=72) (n=75)
- 13 weeks (n=70) (n=75)
- 26 weeks (n=65) (n=75)
- 52 weeks (n=69) (n=75)

*Number of patients measured, # number of patients in analysis
MT= manual therapy, BGA= behavioural graded activity
Is behavioural graded activity more effective than Manual Therapy? – an RCT

Table 1. Baseline characteristics of the two treatment groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>BGA (n=71)</th>
<th>Manual therapy (n=75)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ( mean; SD)</td>
<td>44.5 (12.0)</td>
<td>45.6 (11.1)</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>59.2</td>
<td>62.7</td>
</tr>
<tr>
<td>Complaints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain (%)</td>
<td>93.0</td>
<td>88.0</td>
</tr>
<tr>
<td>Limitation of movement (%)</td>
<td>57.7</td>
<td>56.0</td>
</tr>
<tr>
<td>Headache (%)</td>
<td>63.4</td>
<td>69.3</td>
</tr>
<tr>
<td>Dizziness (%)</td>
<td>36.6</td>
<td>28.0</td>
</tr>
<tr>
<td>Fatigue (%)</td>
<td>56.3</td>
<td>48.0</td>
</tr>
<tr>
<td>Pain intensity (NRS 0-10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain now (mean, SD)</td>
<td>5.5 (2.2)</td>
<td>5.0 (2.1)</td>
</tr>
<tr>
<td>Mean pain previous month (mean, SD)</td>
<td>6.2 (1.5)</td>
<td>6.0 (1.5)</td>
</tr>
<tr>
<td>NDI (mean, SD)</td>
<td>14.7 (6.2)</td>
<td>13.4 (7.4)</td>
</tr>
<tr>
<td>Graded Chronic Pain Scale (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low intensity</td>
<td>11.0</td>
<td>14.7</td>
</tr>
<tr>
<td>high intensity</td>
<td>57.5</td>
<td>58.7</td>
</tr>
<tr>
<td>moderate limiting</td>
<td>15.1</td>
<td>12.0</td>
</tr>
<tr>
<td>severe limiting</td>
<td>16.4</td>
<td>14.7</td>
</tr>
<tr>
<td>SF 36 (mean, SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical component summary</td>
<td>42.5 (7.9)</td>
<td>44.8 (7.3)</td>
</tr>
<tr>
<td>Mental component summary</td>
<td>46.8 (11.4)</td>
<td>47.5 (12.2)</td>
</tr>
</tbody>
</table>

BGA= behavioural graded activity, NRS= Numerical Rating Scale, NDI= Neck Disability Index, SF 36= Short Form 36, SD= Standard Deviation. The baseline characteristics of the Pain Coping and Cognition List, the 4 Dimensional Symptom Questionnaire and the Tampa Scale for Kinesiophobia can be found in Table 3
Chapter 4

Effect of the interventions

Intention-to-treat analyses

Table 2 shows the percentages and odds ratios for the GPE, and the means and regression coefficients (that can be interpreted as mean difference) for the NRS and the NDI at 13 and 52 weeks. At 52 weeks, the BGA group scored slightly better, for the GPE expressed as an OR of 0.76 (0.21-2.68), for the NRS expressed as a regression coefficient or mean difference of 0.99 (0.15-1.83) points, and for the NDI expressed as a mean difference of 2.42 (0.52-4.32) points (Figures 2, 3 and 4). The only statistically significant overall effect was found on the NDI in favour of the BGA treatment. This effect was present at all follow-up moments. For all the other primary outcomes there was no statistically significant difference between the two groups. Considering the results at 13 weeks, the effects of the BGA treatment were achieved earlier than the effects of the MT.

Table 3 shows the effects of the two interventions on the secondary outcome measurements, the PCCL, the 4 DSQ and the TSK. There was no statistically significant overall difference in effect between the two interventions. Only somatisation, a domain within the 4 DSQ, showed a significant difference in favour of the BGA treatment at 52 weeks. There were also no differences in effects between the secondary outcomes SF 36 and GCPS (data not shown here).
Table 2. Multilevel model-based mean scores at baseline, post-treatment and follow-up and the odds ratios and regression coefficients for the primary outcome measures

<table>
<thead>
<tr>
<th>Primary outcome measure</th>
<th>BGA</th>
<th>Manual therapy</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• GPE (%yes)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall effect: $X^2 = 5.26$; 4 df (p=0.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 weeks</td>
<td>85.6%</td>
<td>70.1%</td>
<td>0.39 (0.12 ; 1.28)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>89.4%</td>
<td>86.5%</td>
<td>0.76 (0.21 ; 2.68)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pain (NRS 0-10)</td>
</tr>
<tr>
<td>Overall effect: $X^2 = 5.26$; 4 df (p=0.23)</td>
</tr>
<tr>
<td>baseline</td>
</tr>
<tr>
<td>13 weeks</td>
</tr>
<tr>
<td>52 weeks</td>
</tr>
</tbody>
</table>

| • NDI (0-50)            |     |                |                    |
| Overall effect: $X^2 = 9.66$; 4 df (p=0.05) |     |                |                    |
| baseline                | 14.68 | 13.40          | 2.05 (0.17 ; 3.93) |
| 13 weeks                | 5.55 | 6.28           | (0.52 ; 4.32)      |
| 52 weeks                | 4.28 | 5.42           |                    |

In all multilevel analyse the baseline scores were accounted for. The mean difference can be interpreted as regression coefficient between interventions, GPE= Global Perceived Effect, *model based with random effect= 0, BGA= behavioural graded activity, NRS= Numerical rating Scale, NDI= Neck Disability Index, CI= confidence interval, $X^2$ = Chi-square, df= degrees of freedom.
Chapter 4

Figure 2. Results of the Global Perceived Effect (GPE) during one year follow up

Max. pain

Figure 3. Results of mean pain intensity on the Numerical Rating Scale during one year follow-up
Is behavioural graded activity more effective than Manual Therapy? – an RCT

Figure 4. Results of the Neck Disability Index during one year follow-up

Table 3 shows the effects of both interventions on the secondary outcome measurements, the PCCL, the 4 DSQ and the Tampa scale. There was no statistically significant overall difference in effect between the two interventions. Only somatisation, a domain within the 4 DSQ, showed a significant difference in favour of the BGA programme at 52 weeks. On the secondary outcomes SF 36 and GCPS, data not shown here, the overall effects were the same.

Per protocol analyses
Despite the training course and the follow-up sessions, we found that only 52.1% of the PTs adhered adequately to the BGA programme, according to the strict principles. On the other hand, 80% the manual therapy treatment sessions consisted of manipulations and/or mobilisations with or without exercises. The results of the per protocol analyses (data not shown) were very similar to the results of the intention-to-treat analyses.
Table 3. Multilevel model based mean scores at baseline, post treatment and follow-up and the regression coefficients for the secondary outcome measures.

<table>
<thead>
<tr>
<th>Scales of PCCL</th>
<th>BGA</th>
<th>Manual therapy</th>
<th>Mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Catastrophizing (range 1-6)</td>
<td>2.25</td>
<td>2.35</td>
<td>-0.04 (-0.27 ; 0.19)</td>
</tr>
<tr>
<td>Overall effect: X^2 = 2.53; 4 df (p=0.64)</td>
<td>1.72</td>
<td>1.77</td>
<td>0.06 (-0.17 ; 0.30)</td>
</tr>
<tr>
<td>baseline</td>
<td>1.62</td>
<td>1.78</td>
<td></td>
</tr>
<tr>
<td>13 weeks</td>
<td>2.25</td>
<td>2.35</td>
<td>-0.04 (-0.27 ; 0.19)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>3.30</td>
<td>3.46</td>
<td>-0.08 (-0.36 ; 0.19)</td>
</tr>
<tr>
<td>• Coping (range 1-6)</td>
<td>3.30</td>
<td>3.46</td>
<td>-0.08 (-0.36 ; 0.19)</td>
</tr>
<tr>
<td>Overall effect: X^2 = 5.00; 4 df (p=0.29)</td>
<td>3.34</td>
<td>3.41</td>
<td>-0.28 (-0.56 ; 0.00)</td>
</tr>
<tr>
<td>baseline</td>
<td>3.54</td>
<td>3.42</td>
<td></td>
</tr>
<tr>
<td>13 weeks</td>
<td>3.34</td>
<td>3.41</td>
<td>-0.08 (-0.36 ; 0.19)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>3.54</td>
<td>3.42</td>
<td>-0.28 (-0.56 ; 0.00)</td>
</tr>
<tr>
<td>• Internal pain control (range 1-6)</td>
<td>3.67</td>
<td>3.83</td>
<td>-0.28 (-0.54 ; -0.03)</td>
</tr>
<tr>
<td>Overall effect: X^2 = 6.03; 4 df (p=0.20)</td>
<td>4.10</td>
<td>3.97</td>
<td>-0.18 (-0.44 ; 0.08)</td>
</tr>
<tr>
<td>baseline</td>
<td>4.13</td>
<td>4.11</td>
<td></td>
</tr>
<tr>
<td>13 weeks</td>
<td>4.13</td>
<td>4.11</td>
<td>-0.18 (-0.44 ; 0.08)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>4.13</td>
<td>4.11</td>
<td>-0.18 (-0.44 ; 0.08)</td>
</tr>
<tr>
<td>• External pain control (range 1-6)</td>
<td>2.94</td>
<td>3.21</td>
<td></td>
</tr>
<tr>
<td>Overall effect: X^2 = 5.78; 4 df (p=0.22)</td>
<td>2.72</td>
<td>2.73</td>
<td>0.19 (-0.07 ; 0.45)</td>
</tr>
<tr>
<td>baseline</td>
<td>2.44</td>
<td>2.60</td>
<td>-0.11 (-0.37 ; 0.16)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>2.72</td>
<td>2.73</td>
<td>0.19 (-0.07 ; 0.45)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>2.44</td>
<td>2.60</td>
<td>-0.11 (-0.37 ; 0.16)</td>
</tr>
</tbody>
</table>
In all multilevel analyses the baseline scores were accounted for. Mean difference between interventions can be interpreted as regression coefficient, BGA= behavioural graded activity, PCCL= Pain Coping and cognition List. CI= confidence interval, $X^2$ = Chi-square, df= degrees of freedom.

### Table 3. Continued

<table>
<thead>
<tr>
<th>Scales of PCCL</th>
<th>BGA</th>
<th>Manual therapy</th>
<th>Mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Distress (range 0-32) Overall effect: $X^2= 4.57$; 4 df (p=0.33)</td>
<td>8.37</td>
<td>8.58</td>
<td>0.42 (-1.39 ; 2.23)</td>
</tr>
<tr>
<td>baseline</td>
<td>4.79</td>
<td>5.41</td>
<td>-</td>
</tr>
<tr>
<td>13 weeks</td>
<td>3.50</td>
<td>5.03</td>
<td>1.32 (-0.52 ; 3.15)</td>
</tr>
<tr>
<td>52 weeks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Depression (range 0-12) Overall effect: $X^2= 2.04$; 4 df (p=0.73)</td>
<td>0.49</td>
<td>0.67</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>0.27</td>
<td>0.24</td>
<td>0.22 (-0.78 ; 0.34)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>0.24</td>
<td>0.43</td>
<td>0.01 (-0.56 ; 0.57)</td>
</tr>
<tr>
<td>52 weeks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Fear (range 0-24) Overall effect: $X^2= 8.26$; 4 df (p=0.08)</td>
<td>1.07</td>
<td>1.97</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>0.52</td>
<td>1.30</td>
<td>0.11 (-0.89 ; 0.67)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>0.51</td>
<td>1.07</td>
<td>0.34 (-1.13 ; 0.45)</td>
</tr>
<tr>
<td>52 weeks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Somatisation (range 0-32) Overall effect: $X^2= 2.53$; 4 df (p=0.64)</td>
<td>9.47</td>
<td>9.86</td>
<td>-</td>
</tr>
<tr>
<td>baseline</td>
<td>5.31</td>
<td>5.90</td>
<td>0.20 (-1.23 ; 1.63)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>4.27</td>
<td>5.71</td>
<td>1.04 (0.81 ; 1.28)</td>
</tr>
</tbody>
</table>
In all multilevel analyse the baseline scores were accounted for. Mean difference between interventions can be interpreted as regression coefficient, BGA= behavioural graded activity, TSK= Tampa Scale for Kinesiophobia, CI= confidence interval, $X^2$ = Chi-square, df= degrees of freedom

### Discussion

We compared the effectiveness of a BGA programme to MT for the treatment of sub-acute neck pain. In previous studies MT has been found to be the most appropriate therapy. It was concluded that on the primary outcome measures there was a marginal difference of effect in favour of the BGA treatment, which only reached statistical significance on the NDI. Moreover, the effect of a BGA treatment was achieved earlier than the effect of the MT. We found no significant difference in effect on the secondary outcomes, which included mainly psychological measures.

The minimal clinically important change has been estimated to be at least 3.5 points on the NDI, and at least 2.5 points on the NRS. In the current trial the change scores on the NDI were 10.4 points for the BGA group and 8.0 points for the MT group. On the NRS these mean changes were 4.4 and 3.5, respectively. This implies that in both intervention groups substantial improvements were observed, but that the differences in improvements between the two groups were small. It can be questioned to what extent the improvements in the two intervention groups are due to natural recovery from sub-acute neck pain. In a previous trial carried out by Hoving et al, MT appeared to be clearly more effective than usual care (mainly based on a ‘wait and see’ policy) provided by GPs for patients with sub-acute and chronic neck pain. In

<table>
<thead>
<tr>
<th>TSK (range 17-65)</th>
<th>BGA</th>
<th>Manual Therapy</th>
<th>Mean Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall effect: 4 df (p=0.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline</td>
<td>39.57</td>
<td>37.94</td>
<td>-1.35 (-4.06 ; 1.39)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>38.57</td>
<td>35.61</td>
<td>-0.23 (-2.78 ; 3.25)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>39.33</td>
<td>37.93</td>
<td></td>
</tr>
</tbody>
</table>
the current trial a BGA treatment was found to be marginally more effective than MT. Therefore, we conclude that both interventions are effective, and that the natural course is not fully responsible for the improvement achieved in the two interventions.

Due to the focus on improving functional status and a lack of focus on pain reduction, a significant difference in functional status (NDI) was in line with the expectations. Surprisingly, no differences in effects were found on kinesiophobia, coping styles or other psychological variables. There might be several explanations why the BGA treatment was not more convincingly beneficial than MT, as we had anticipated:

1) One reason could be that the hypothesized working mechanism of a BGA programme was not effective for patients with sub-acute neck pain. Improvement in psychological outcomes such as fear of movement, distress, coping, depression, etc., was to be expected, thereby resulting in an increase level of function. However, after the intervention there was no significant change in these psychological variables. This could be due to the mainly low initial scores for these psychological outcomes at baseline: very few patients scored ‘moderate’ to ‘high’ on the 4 DSQ, PCCL or TSK, so no great improvement could be expected.

2) Another explanation might concern the attitudes and beliefs of the PTs who provided the BGA treatment. The training course for the PTs prior to the intervention was aimed, among other things, at changing their attitude from a biomedical attitude to a biopsychosocial attitude. We regard this biopsychosocial attitude as vital for the optimal provision of a BGA treatment. Since the strict protocol of the BGA programme, which was based on a hands-off strategy, was not always adequately adhered to, we question the presence of a biopsychosocial attitude in PTs, even if they had already followed previous courses. The principles of BGA are not easily learned or applied, as was also obvious in GP care\(^{17;18}\). A two-day training course is perhaps insufficient for the PTs to adopt a sufficiently strong biopsychosocial attitude, and this may have hampered optimal provision of the therapy.

3) Yet another explanation might concern insufficient flexibility of daily practice in primary care. A BGA programme requires a rather strict protocol in terms of organisation and frequency of treatment sessions. It turned out to be difficult to adhere to a strict protocol, and this might have resulted in a sub-optimal provision of the BGA treatment in this trial.

4) Another factor which can influence the effectiveness of BGA treatment on neck pain is the duration of the complaints. In this trial we included patients with sub-acute neck pain. Consequently, many
patients had a quick recovery, so the pre-determined treatment aims were achieved sooner than was expected. As a consequence, the intensity of the programme, as determined at baseline, was too low for some patients, and therefore compliance with the strict protocol was also rather low.

Although we initially aimed to include 180 participants we were only able to include 146 patients. The GPs were responsible for the recruitment of patients, but despite the fact that each of the participating GPs (n=72) who were contacted individually by the researcher agreed to participate, only 49% actually referred patients to the study. There was also a change in legislation during the recruitment period which de-listed physical therapy from basic health care insurance, and from January 2005 onwards patients needed an additional health care insurance to cover the costs of physical therapy and MT. GPs were therefore more reluctant than before to refer patients to a physical therapist or a manual therapist. Considering the marginal differences in effectiveness between the interventions, it seems unlikely that there would have been any drastic change in results with the inclusion of an additional 34 patients.

We therefore consider that the number of patients included in this study was sufficient to draw conclusions.

In the Hoving trial\textsuperscript{16} MT appeared to be more effective than usual care or physical therapy for patients with sub-acute and chronic neck pain, but in our study the BGA treatment appeared to be slightly more effective than MT. Based on the results of both studies, we conclude that patients with sub-acute non-specific neck pain can benefit from both interventions, although the BGA treatment is slightly more effective than the MT, and the effect of the BGA treatment is achieved more rapidly. However, although referral for BGA treatment is clearly the best choice, GPs still do not know to whom they should refer a patient with sub-acute neck pain, because it is not self-evident that a PT in a primary care setting will conduct a BGA programme in an optimal fashion. PTs need additional education in order to equip them for adherence to a BGA programme.

In the current situation a GP might refer a patient to a PT who is trained to provide BGA treatment if the GP expect that the patient has fear of movement or other psychological problems. Other patients would be referred for MT.

In future studies that aim to study the effect of a BGA programme, it seems reasonable to suggest that the focus should be more on patients...
with psychological characteristics, such as high scores for fear of movement, distress, depression, etc. Those patients might benefit most from a BGA programme.
Reference List


Is behavioural graded activity more effective than Manual Therapy? – an RCT


Is behavioural graded activity more effective than Manual Therapy? – an RCT


Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain?

Submitted as; Jan J.M. Pool, Raymond W.J.G. Ostelo, Dirk Knol, Lex M. Bouter, Henrica C.W. de Vet; Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain?
Chapter 5

Abstract

Our aim was to identify psychological factors that are prognostic indicators of the short and long term outcome of patients with sub-acute neck pain in terms of perceived recovery, pain and functional disability. An increasing amount of attention is being paid to psychological factors in the development and maintenance of pain and disability. Furthermore, psychological and social factors are possibly involved in the transition from acute to chronic neck pain. This prospective study was conducted within the framework of a randomised clinical trial comparing two types of conservative therapy in 146 patients with non-specific sub-acute neck pain.

Multilevel techniques were used for data-analysis. The short and long term results for the three outcomes were very diverse. Furthermore, the explained variance in the short term ranged from 16% to 30%, and from 6% to 34% in the long term. This can be considered to be low, and implies that it is difficult to predict the course of neck pain. The sub-scales of the used questionnaires, i.e. the Pain Coping and Cognition List, and the 4 Dimensional Symptom Questionnaire, did not contribute significantly to all of the multilevel models. Only the factor ‘fear of movement’ (Tampa scale) was consistently and significantly present in the univariable analysis for all outcomes at both follow-up measurements, and also for the short term outcome pain and disability in the multivariable analysis. We conclude that we were unable to identify a core set of prognostic psychological factors that predict the short and long term outcome of sub-acute neck pain. Further prognostic research is needed to achieve more consistent results.
Introduction

Neck pain is a common musculoskeletal disorder. The point prevalence for neck pain in the general population varies between 9.5% and 22.0%\(^{4,20}\). Each year approximately one-third of all adults will experience neck pain\(^8\). The Quebec Task Force\(^1\) discriminates between the acute phase (0-6 weeks), the sub-acute phase (6-12 weeks) and the chronic phase (longer than 12 weeks). Some 5-10% of the neck complaints will develop into chronic neck pain. The main feature of neck pain is pain in the cervical region, often accompanied by restriction in the range of motion. This leads to functional limitations, for instance when looking over the shoulder or working with a computer\(^2\). The pain can arise from many structures in the cervical region, especially the spine and soft tissues, but there are no data on the prevalence of particular causes of acute or chronic neck pain\(^3\) and there are no valid means in clinical practice to distinguish one suggested cause of the pain from another. Therefore, the most accurate diagnosis in most cases is asymptomatic or non-specific neck pain\(^3\). Risk factors for the occurrence of neck pain are physical load factors, such as vibration, flexion of the neck, bad sitting posture, and heavy lifting\(^2\). Furthermore, psychological and social factors are also reported to aggravate and perpetuate neck pain, and are believed to be related to the onset of acute neck pain and the transition into chronic neck pain\(^2,16\). There is increasing evidence that psychological and social factors can influence the course of pain, and can play an important role in the development of chronic musculoskeletal disorders\(^5,18,24\). Identifying factors that predict the clinical course of sub-acute neck pain might therefore give indications for the further development of effective treatment strategies.

In primary care, some prognostic factors are routinely included in history-taking, for example high pain levels, and a previous history of neck pain\(^8,14\), but a structural search for psychological and social factors is not common practice. Factors such as the attitudes and beliefs of the patients, coping, depression, psychological distress, illness behaviour and anxiety are all factors which, according to the bio-psychosocial model, can influence the course and experience of pain\(^10,16\). Although some research has focussed on prognostic factors, complete assessments of all potential psychological factors are rare. Therefore, our objective was to identify psychological prognostic indicators of the short and long-term outcome of sub-acute neck pain in terms of perceived recovery, pain and functional disability.
Chapter 5

We carried out a prognostic study, taking into account the variability of the practitioners: manual therapists and physical therapists, because inter-practitioner variability can be substantial, due to practice organisation, professional norms, therapist style, and background, etc.

Materials and methods

Our prognostic study was conducted within the framework of a randomised clinical trial on the effectiveness of manual therapy compared to a graded activity program performed by physical therapists, on patients with sub-acute neck pain\textsuperscript{21}. At baseline, 146 patients completed a questionnaire containing questions about potential prognostic indicators such as: gender, age, history of neck complaints, severity of the pain (table 1). Furthermore, the 4 dimensional symptom questionnaire (4DSQ)\textsuperscript{28} was used to measure: somatisation, distress, depression and fear. The 4DSQ appears to be a valid questionnaire, with acceptable reliability \textsuperscript{30}, and the Pain Coping and Cognition List PCCL\textsuperscript{25} was used to measure: catastrophising, coping, and internal and external pain control. The PCCL is based on a compilation of the Pain Coping List, the Pain Control List and the Coping and Pain Questionnaire, the internal consistency seems to be good,( Cronbach’s α between 0.80 and 0.85), the test-retest reliability was moderate to good, (r between 0.64 and 0.79) and a fair construct validity\textsuperscript{9}. Fear of movement was measured with the Tampa Scale for Kinesiophobia (TSK)\textsuperscript{15}, the TSK showed a good internal consistency and substantial test-retest reliability\textsuperscript{26}. The level of chronicity was assessed with the Graded Chronic pain Scale (GCPS)\textsuperscript{32}. The patients preference or non-preference for therapy, manual therapy, physical therapy, and general practitioner attitude towards neck pain were assessed with the Pain Beliefs and Attitude Scale\textsuperscript{19}.

Three primary outcome measurements were defined and measured at 12 and 52 weeks.

1) The patients rated their perceived recovery on a 7-point ordinal rating scale, ranging from ‘completely recovered’ to ‘worse than ever’. Recovery was a priori defined as ‘completely recovered’ or ‘much improved’, as reported by the patient.

2) The severity of neck pain was scored on an 11-point Numerical Rating Scale (NRS). Recovery from pain was a priori defined as an NRS score of ≤ 1.

3) The Neck Disability Index (NDI) was used to measure functional status\textsuperscript{31}.
**Statistical analysis**

The relationship between each potential prognostic indicator and outcome was evaluated, adjusting for the randomly allocated intervention. As the interventions were performed by a number of therapists, we took patients clusters under therapists into account in the analysis. Therefore a multilevel analysis was performed, with two levels: patients and therapists. For continuous data, i.e. from the NDI, a linear regression model was fitted in a four-step strategy, applying the Likelihood ratio test with a significant level of 10\%\(^6\). The four steps are:

1) The univariable step, in which the \(-2\) log Likelihood (\(-2\text{LL}\)) is compared with the null model, which consists of the intercept and therapy, to determine which variable significantly reduces the value of this statistic.

2) These variables are then included in a multivariable model, and variables which do not significantly increase the value of the \(-2\text{LL}\) when they are omitted from the model are removed.

3) Variables which were not important in the first step may become important in the presence of others, these variables are added to the model from step 2, one at a time, to see if there is any significant reduction of the \(-2\text{LL}\).

4) A final check is made to ensure that none of the variables in the model can be omitted without significantly increasing the value of \(-2\text{LL}\).

For dichotomous outcomes such as the GPE and the NRS we made a logistic regression model in which the odds ratio (OR) and 95\% confidence interval (CI) were calculated. The same 4-step strategy was adhered to, but instead using the likelihood ratio test, we used the Wald-statistic test. Again, a level of significance of 10\% was set for the model strategy. All analysis were performed in MLWin 2.02.

**Results**

Between January 2003 and December 2004, 146 patients were included in the trial. At 52 weeks follow-up 18 patients had dropped out, but 8 patients had provided information by phone about their perceived recovery. Therefore, the analysis included 146 eligible patients at 12 weeks, and 128 eligible patients at 52 weeks, with the exception of the GPE outcome which included the scores of 136 patients.
### Table 1. Baseline characteristics of the two treatment groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>BGA (n=71)</th>
<th>Manual therapy (n=75)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (mean; SD)</strong></td>
<td>44.5 (12.0)</td>
<td>45.6 (11.1)</td>
</tr>
<tr>
<td><strong>Gender (% female)</strong></td>
<td>59.2</td>
<td>62.7</td>
</tr>
<tr>
<td><strong>Complaints</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain (%)</td>
<td>93</td>
<td>88</td>
</tr>
<tr>
<td>Limitation of movement (%)</td>
<td>57.7</td>
<td>56.0</td>
</tr>
<tr>
<td>Headache (%)</td>
<td>63.4</td>
<td>69.3</td>
</tr>
<tr>
<td>Dizziness (%)</td>
<td>36.6</td>
<td>28.0</td>
</tr>
<tr>
<td>Fatigue (%)</td>
<td>56.3</td>
<td>48.0</td>
</tr>
<tr>
<td><strong>Pain intensity (NRS 0-10)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain now (mean, SD)</td>
<td>5.5 (2.2)</td>
<td>5.0 (2.1)</td>
</tr>
<tr>
<td>Mean pain last month (mean, SD)</td>
<td>6.2 (1.5)</td>
<td>6.0 (1.5)</td>
</tr>
<tr>
<td><strong>NDI (mean, SD)</strong></td>
<td>14.7 (6.2)</td>
<td>13.4 (7.4)</td>
</tr>
<tr>
<td><strong>Graded Chronic Pain Scale (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low intensity</td>
<td>11.0</td>
<td>14.7</td>
</tr>
<tr>
<td>high intensity</td>
<td>57.5</td>
<td>58.7</td>
</tr>
<tr>
<td>moderate limiting</td>
<td>15.1</td>
<td>12.0</td>
</tr>
<tr>
<td>severe limiting</td>
<td>16.4</td>
<td>14.7</td>
</tr>
<tr>
<td><strong>Tampa (mean, SD)</strong></td>
<td>32.2 (6.4)</td>
<td>32.3 (5.7)</td>
</tr>
<tr>
<td><strong>4 DSQ (mean, SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distress</td>
<td>8.7 (6.2)</td>
<td>8.6 (7.6)</td>
</tr>
<tr>
<td>depression</td>
<td>0.6 (1.4)</td>
<td>0.6 (1.7)</td>
</tr>
<tr>
<td>fear</td>
<td>1.3 (2.5)</td>
<td>2.0 (3.6)</td>
</tr>
<tr>
<td>somatisation</td>
<td>9.6 (3.8)</td>
<td>9.7 (5.1)</td>
</tr>
<tr>
<td><strong>SF 36 (mean, SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical component summary</td>
<td>42.5 (7.9)</td>
<td>44.8 (7.3)</td>
</tr>
<tr>
<td>Mental component summary</td>
<td>46.8 (11.4)</td>
<td>47.5 (12.2)</td>
</tr>
<tr>
<td><strong>Pain Coping Cognition List (mean, SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catastrofying</td>
<td>2.3 (0.8)</td>
<td>2.3 (1.0)</td>
</tr>
<tr>
<td>Coping</td>
<td>3.3 (0.9)</td>
<td>3.5 (0.9)</td>
</tr>
<tr>
<td>Internal pain control</td>
<td>3.7 (0.9)</td>
<td>3.8 (0.9)</td>
</tr>
<tr>
<td>External pain control</td>
<td>3.0 (0.9)</td>
<td>3.2 (0.9)</td>
</tr>
</tbody>
</table>

SD = Standard Deviation, NDI = Neck Disability Index, 4 DSQ = Four-Dimensional Symptom Questionnaire, SF 36 = Short Form 36


**Table 2.** Recovery, pain and disability among sub-acute neck pain patients at 12 and 52 week.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>12 weeks</th>
<th>52 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived recovery (%)</td>
<td>70.5 %</td>
<td>77.2 %</td>
</tr>
<tr>
<td>NRS score ≤ 1 (%)</td>
<td>48.6 %</td>
<td>68.2 %</td>
</tr>
<tr>
<td>NDI (0-50) (mean and SD)</td>
<td>5.7 (5.4)</td>
<td>4.5 (5.2)</td>
</tr>
</tbody>
</table>

NRS = Numerical Rating Scale for pain, NDI = Neck Disability Index
SD = standard deviation

**Perceived recovery**

Table 3 presents the prognostic indicators of perceived recovery at 12 weeks. The univariable analysis showed a significant association with a worse outcome for greater fear of movement, gender (male), and headache. Preference for physical therapy predicted a more favourable outcome, although this was not statistically significant. In the multivariable analysis, only headache and preference for physical therapy were significant predictors of outcome. The explained variance in the model was 17%

The univariable analysis at 52 weeks (not presented) showed an association only for fear of movement, OR=0.92 (CI=0.87-0.99), greater fear of movement being associated with a worse outcome. The explained variance ($R^2$) in this model was 6%.

**Pain**

Table 4 presents the prognostic indicators of pain at 12 weeks, and the results of univariable analysis show the contribution of catastrophising, fear of movement, somatisation, fear, gender, headache and severity of complaints. In the multivariable model, fear of movement, gender and severity of complaints were significant predictors. Greater fear of movement and gender (male) were associated with a worse outcome, and a higher score for severity of complaints was associated with a more favourable outcome. The explained variance in this model was 16%. The univariable analysis at 52 weeks (not presented) showed a significant association only of distress, a higher level of distress being associated with a more favourable outcome, with an explained variance ($R^2$) of 6%.

**NDI**

Table 5 presents the prognostic indicators of the NDI at 12 weeks. In the univariable analysis, all indicators except coping, severity of
complaints and patient preference were significantly associated with the score at 12 weeks. In the multivariable model, fear of movement, somatisation, gender (male), age and the score of the GCPS, especially ‘moderately limiting’ compared to ‘low intensity of complaints’ were associated with a worse outcome and internal pain control was associated with a more favourable outcome. The explained variance ($R^2$) in this model was 30%.

The univariable analysis at 52 weeks (not presented) showed an association for catastrophysing, fear of movement, somatisation, fear, distress, gender, headache, and the GCPS and NDI scores at baseline. After inclusion in a multivariable model, only higher GCPS scores and the baseline NDI score were associated with the a worse outcome. The explained variance ($R^2$) in this model is 34%. In Table 6 we summarised our findings.
Table 3. Results of the univariable and multivariable multilevel analysis of perceived recovery at 12 weeks (n=146)

<table>
<thead>
<tr>
<th></th>
<th>univariable</th>
<th></th>
<th>multivariable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>OR (95% CI)</td>
<td>B (SE)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>PCCL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catastrophising</td>
<td>-0.562 (0.485)</td>
<td>0.57 (0.22 - 1.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coping</td>
<td>0.197 (0.211)</td>
<td>1.22 (0.81 - 1.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>internal pain control</td>
<td>0.180 (0.207)</td>
<td>1.20 (0.80 - 1.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external pain control</td>
<td>0.081 (0.038)</td>
<td>1.08 (0.73 - 1.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSK</td>
<td>-0.051 (0.030)</td>
<td>0.95 (0.90 - 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>somatisation</td>
<td>-0.032 (0.040)</td>
<td>0.96 (0.90 - 1.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fear</td>
<td>-0.237 (0.225)</td>
<td>0.79 (0.51 - 1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distress</td>
<td>-0.070 (0.205)</td>
<td>0.93 (0.62 - 1.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>-0.290 (0.321)</td>
<td>0.75 (0.40 - 1.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (♂)</td>
<td>-0.686 (0.369)</td>
<td>0.50 (0.23 - 0.91)</td>
<td>-1.303 (0.467)</td>
<td>0.27 (0.11 - 0.68)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.011 (0.016)</td>
<td>0.99 (0.96 - 1.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>-1.245 (0.461)</td>
<td>0.29 (0.12 - 0.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity of complaints</td>
<td>0.049 (0.127)</td>
<td>1.05 (0.82 - 1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of neck complaints</td>
<td>-0.444 (0.374)</td>
<td>0.64 (0.31 - 1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pt</td>
<td>1.399 (0.783)</td>
<td>4.05 (0.87 - 18.8)</td>
<td>1.525 (0.794)</td>
<td>4.60 (0.97 - 21.79)</td>
</tr>
<tr>
<td>mt</td>
<td>0.545 (0.464)</td>
<td>1.72 (0.69 - 1.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCPS (grade)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= low intensity (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2= high intensity</td>
<td>0.585 (0.584)</td>
<td>1.79 (0.57 - 5.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3= moderately limiting</td>
<td>-0.489 (0.689)</td>
<td>0.61 (0.16 - 2.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= severely limiting</td>
<td>0.262 (0.696)</td>
<td>1.30 (0.33 - 5.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>purely biomedical (ref)</td>
<td></td>
<td></td>
<td></td>
<td>explained variance R² = 17%</td>
</tr>
<tr>
<td>more biomedical</td>
<td>-0.755 (0.529)</td>
<td>0.47 (0.17 - 1.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td>0.772 (0.550)</td>
<td>2.16 (0.74 - 6.36)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Psychological factors in patients with sub-acute neck pain
<table>
<thead>
<tr>
<th></th>
<th>univariable</th>
<th></th>
<th>multivariable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>OR (95% CI)</td>
<td>B (SE)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td><strong>PCCL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catastrofyzing</td>
<td>-0.862 (0.452)</td>
<td>0.42 (0.17 – 1.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coping</td>
<td>-0.119 (0.190)</td>
<td>0.89 (0.60 – 1.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>internal pain</td>
<td>0.020 (0.188)</td>
<td>1.22 (0.71 – 1.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>control</td>
<td>-0.035 (0.180)</td>
<td>0.97 (0.68 – 1.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>external pain</td>
<td>-0.059 (0.029)</td>
<td>0.94 (0.89 – 1.00)</td>
<td>-0.079 (0.031)</td>
<td>0.92 (0.87 – 0.98)</td>
</tr>
<tr>
<td>control</td>
<td>TSK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.059 (0.029)</td>
<td>0.94 (0.89 – 1.00)</td>
<td>-0.079 (0.031)</td>
<td>0.92 (0.87 – 0.98)</td>
<td></td>
</tr>
<tr>
<td>4DSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>somatisation</td>
<td>-0.085 (0.040)</td>
<td>0.92 (0.85 – 0.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fear</td>
<td>-0.561 (0.230)</td>
<td>0.57 (0.36 – 0.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distress</td>
<td>-0.322 (0.198)</td>
<td>0.72 (0.50 – 1.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>-0.418 (0.320)</td>
<td>0.66 (0.35 – 1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (♂)</td>
<td>-0.967 (0.350)</td>
<td>0.38 (0.19 – 0.76)</td>
<td>-1.147 (0.374)</td>
<td>0.32 (0.15 – 0.66)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.000 (0.014)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>-0.496 (0.355)</td>
<td>0.60 (0.30 – 1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity of complaints</td>
<td>0.255 (0.124)</td>
<td>1.29 (1.10 – 1.65)</td>
<td>0.299 (0.130)</td>
<td>1.35 (1.05 – 1.74)</td>
</tr>
<tr>
<td>History of neck complaints</td>
<td>-0.031 (0.335)</td>
<td>0.97 (0.48 – 1.94)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient preference</td>
<td>none (ref)</td>
<td>pt 0.845 (0.543)</td>
<td>2.33 (0.81 – 6.70)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mt 0.151 (0.407)</td>
<td>1.16 (0.80 – 2.58)</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. continued

<table>
<thead>
<tr>
<th></th>
<th>univariable</th>
<th>multivariable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b (SE)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>GCPS (grade)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= low intensity (ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = high intensity</td>
<td>0.284 (0.554)</td>
<td>1.33 (0.45 – 3.93)</td>
</tr>
<tr>
<td>3 = moderately limiting</td>
<td>-0.043 (0.671)</td>
<td>0.96 (0.26 – 3.59)</td>
</tr>
<tr>
<td>4 = severely limiting</td>
<td>-0.179 (0.660)</td>
<td>0.84 (0.23 – 3.05)</td>
</tr>
<tr>
<td>GP attitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>purely biomedical (ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>more biomedical</td>
<td>0.400 (0.520)</td>
<td>1.49 (0.54 – 4.13)</td>
</tr>
<tr>
<td>neutral</td>
<td>0.784 (0.490)</td>
<td>2.19 (0.89 – 5.72)</td>
</tr>
<tr>
<td>Pain at baseline</td>
<td>-0.096 (0.079)</td>
<td>0.91 (0.84 – 5.72)</td>
</tr>
</tbody>
</table>

Table 3 and Table 4: SE=standard error, OR=odds ratio CI=95% confidence interval, b=regression-coefficient, ref= reference category, pt=physical therapy, mt= manual therapy, GP= general practitioner, PCCL= Pain Coping and Cognition List, 4DSQ= 4 dimensional symptom questionnaire, GCPS=Graded Chronic Pain Scale, TSK= Tampa Scale of Kinesiophobia

*two levels: patient and therapists, adjusting for therapy
Table 5. Results of the univariable and multivariable multilevel analysis* of the NDI at 12 weeks.

<table>
<thead>
<tr>
<th></th>
<th>univariable</th>
<th></th>
<th>multivariable</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b(SE)</td>
<td>-2LL</td>
<td>b(SE)</td>
<td>-2LL (model)</td>
</tr>
<tr>
<td>Therapy (mt)</td>
<td>0.0515 (0.872)</td>
<td>899.552</td>
<td>3.311 (1.116)</td>
<td>891.904</td>
</tr>
<tr>
<td>PCCL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catastrofyzing coping</td>
<td>0.106 (0.497)</td>
<td>899.507</td>
<td>-0.884 (0.486)</td>
<td>896.285</td>
</tr>
<tr>
<td>internal pain control</td>
<td>-0.884 (0.486)</td>
<td>896.285</td>
<td>1.011 (0.467)</td>
<td>894.944</td>
</tr>
<tr>
<td>external pain control</td>
<td>1.011 (0.467)</td>
<td>894.944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSK</td>
<td>0.198 (0.070)</td>
<td>891.761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>somatisation</td>
<td>0.311 (0.094)</td>
<td>888.722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fear</td>
<td>1.691 (0.537)</td>
<td>889.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distress</td>
<td>1.332 (0.475)</td>
<td>891.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td>1.709 (0.789)</td>
<td>894.934</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (♂)</td>
<td>3.611 (0.859)</td>
<td>887.755</td>
<td>0.174 (0.089)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.078 (0.037)</td>
<td>895.275</td>
<td>0.077 (0.033)</td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>2.618 (0.899)</td>
<td>891.317</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severity of complaints</td>
<td>0.112 (0.316)</td>
<td>899.419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pt</td>
<td>0.036 (1.346)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mt</td>
<td>-1.732 (1.056)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Continued

<table>
<thead>
<tr>
<th></th>
<th>univariable</th>
<th></th>
<th>multivariable</th>
<th></th>
<th>-2LL</th>
<th></th>
<th>-2LL (model)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b(SE)</td>
<td>-2LL</td>
<td>b(SE)</td>
<td>-2LL</td>
<td>(model)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCPS (grade)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1= low intensity (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = high intensity</td>
<td>-0.048 (1.397)</td>
<td>858.778</td>
<td>-0.954 (1.234)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 = moderately limiting</td>
<td>3.648 (1.691)</td>
<td></td>
<td>3.189 (1.509)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4= severely limiting</td>
<td>1.656 (1.656)</td>
<td></td>
<td>-0.072 (1.548)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP attitude(n=125)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>purely biomedical (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>more biomedical</td>
<td>-0.275 (1.355)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td>-2.150 (1.276)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI-baseline</td>
<td>0.230 (0.061)</td>
<td>885.519</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b=regression-coefficient, SE=standard error, -2LL= -2 Log Likelyhood, ref= reference category, pt=physical therapy, mt= manual therapy, GP= general practitioner, PCCL= Pain Coping and Cognition List, 4DSQ= 4 dimensional symptom questionnaire, GCPS=Graded Chronic Pain Scale, TSK= Tampa Scale of Kinesiophobia
*two levels: patient and therapists, adjusting for therapy.
<table>
<thead>
<tr>
<th>variable</th>
<th>12 weeks</th>
<th></th>
<th></th>
<th>52 weeks</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDI</td>
<td>GPE</td>
<td>Pain</td>
<td>NDI</td>
<td>GPE</td>
<td>Pain</td>
</tr>
<tr>
<td>PCCL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catastrophizing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>internal pain control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>external pain control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DSQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>somatisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>headache</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>severity of complaints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>history of neck complaints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>patient preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCPS (grade)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI baseline score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GPE= Global Perceived Effect, NDI= Neck Disability Index, PCCL= Pain Coping and Cognition List, 4DSQ= 4 dimensional symptom questionnaire, GCPS= Graded Chronic Pain Scale, TSK= Tampa Scale of Kinesiophobia, GP= general practitioner.
Discussion

Our objective was to identify psychological prognostic indicators of the short and long term outcome of sub-acute neck pain in terms of perceived recovery, pain and functional disability. The results of this study show a diverse picture. The only factor which was more or less consistently present in the univariable analysis for all measurements, except for pain at 52 weeks, was fear of movement (TSK). The score on the GCPS scale and especially the factor ‘highly disabling-pain with moderate activity limitations’ for the NDI outcome was a significant prognostic factor in the multivariable analysis at both 12 weeks and 52 weeks. The explained variance in these both models was 30% and 34% which can be considered to be reasonable. The categories of the PCCL and the 4 DSQ contribute very little at both measurement points for all outcomes. This could be due to the fact that the mean scores for the different psychological domains of these questionnaires fluctuate around normal, so there are no patients with extremely high or extremely low scores. A score of ≥ 3.5 in each category of the PCCL (range 1-6) is considered to be high. For catastrophising the mean value was 2.3 (0.9), and 11% of the patients had a high score; for coping the mean was 3.4 (0.9), and 56% had a high score; for internal pain control the mean was 3.8 (0.9), and 41% had a high score; for external pain control the mean was 3.1 (0.9), and 21% had a high score. This lead to the conclusion that only internal pain control at baseline can be considered as substantial. This factor was found to be significantly associated in a multivariable analysis of the NDI at 12 weeks.

The cut-off point for the 4DSQ varies per dimension but, taking into account the level “more severe than normal”, 35% of the patients had such a score on the dimension distress, with a mean score of 8.6 (6.9) and a cut-off point of 10, 20% had a mean score of 0.6 (1.6) for depression with a cut-off point of 2, 4% had a mean score of 1.7 (3.2) for fear with a cut-off point of 8 and 36% had a mean score of 9.7 (4.5) for somatisation with a cut-off point of 10. This can be considered as low scores for psychological factors. Unfortunately, this resulted in very small sub-groups of patients with extreme scores, so too little contrast was found in these psychological domains. However, it reflects the situation in clinical practice.

Another limitation of the current study is that it mainly focussed on psychological factors and not on physical factors.

Comparing our findings with prognostic factors reported in the literature, we noticed that our diverse picture is no exception. Although the patient...
population in the present study was almost the same as in the Hoving study\textsuperscript{13}, factors such as age and previous history of neck pain were not important factors in our study although we did include the same general practitioners and the same location, but different domains of interests, and sub-acute neck pain instead of neck pain (15\% of which was sub-acute).

Linton\textsuperscript{16} found, in an extensive review, that psychological factors were related to neck pain and back pain from onset to the chronic phase. Furthermore, psychological factors were found to be pivotal in the transition from acute to chronic pain, as well as influential in the onset of pain. Based on the results of that review, we hypothesized that these factors can influence the course of neck pain and the outcome of treatment strategies over time. However, Linton did not use any methods of quality rating to assess the articles in his review, and furthermore, in the primary care setting only two of the studies included neck pain, but this were not analyzed separately from low back pain. Whether or not psychological factors predict a poor outcome for sub-acute neck pain can therefore not be concluded from this review. The results of more recent studies\textsuperscript{23,27} on acute low back pain patients also only seem to add to the confusion, and contradictory findings are reported for fear of movement. In the present study we used the TSK, which was developed for the assessment of low back pain. It can be questioned whether fear of movement is the same for patients with back pain as for patients with neck pain, or is fear of movement more domain-specific? Nevertheless, most of the psychological factors investigated in the present study were not consistent in their prediction of outcomes such as perceived recovery, pain and disability.

In the literature the main field of interest is either whiplash-associated disorders (WAD) or neck pain as a separate entity. Factors associated with poor recovery in WAD patients are high initial pain intensity, age, gender and high acute psychological responses\textsuperscript{7,22}. However, Hendriks et al\textsuperscript{11} reported that care-providers could easily identify patients who were at risk for poor recovery with a simple visual analogue scale for initial pain intensity and work-related activities. The traumatic event that precipitates the onset of WAD may have different psychological consequences, and for that reason hard generalisation of the results to other neck pain conditions is difficult.

In patients with neck pain, Hill found an increased risk of persistent neck pain associated with age, comorbid low back pain and cycling\textsuperscript{12}, while Bot et al\textsuperscript{5} found that psychosocial factors, such as passive coping and
fear avoidance, also predicted the outcome of neck and shoulder symptoms. A summary of prognostic factors identified in other studies is presented in Table 7.

In conclusion much attention is paid to psychological factors in the development or maintenance of pain and disability. This paradigm shift from a biomedical approach to a more bio-psychosocial approach is evident in clinical practice, but there is no core set of prognostic psychological factors that predict the outcome of neck pain over time, as is confirmed in the present study. This is probably due to differences in study populations, study settings and definitions of outcome. Furthermore, it is hard to identify consistent psychological prognostic factors with the variety of questionnaires that are used, and the minimal contrast in the psychological variables that was found in the present study. In clinical practice it is thought that sub-acute neck pain in patients with for instance, passive coping and fear of movement can become chronic, but it is still difficult to underpin this seemingly evident statement with scientific evidence. In clinical practice, understanding of the clinical course of neck pain is of importance in decision-making concerning the management of patients with neck pain. Further prognostic research is needed to achieve more consistent results.
Table 7. Summary of prognostic factors for neck pain

<table>
<thead>
<tr>
<th>Author</th>
<th>Population</th>
<th>Follow up months</th>
<th>Psychological factors</th>
<th>Physical factors</th>
<th>General factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leclerc</td>
<td>568 workers</td>
<td>12</td>
<td>• distress</td>
<td></td>
<td>• age</td>
</tr>
<tr>
<td>(1999)</td>
<td></td>
<td></td>
<td>• psycho-somatic problems</td>
<td></td>
<td>• gender</td>
</tr>
<tr>
<td>Bot</td>
<td>443 GP</td>
<td>12</td>
<td>• fear avoidance</td>
<td></td>
<td>• history</td>
</tr>
<tr>
<td>(2004)</td>
<td>1359 gen.pop.‡</td>
<td></td>
<td></td>
<td></td>
<td>• general health</td>
</tr>
<tr>
<td>Hill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• age</td>
</tr>
<tr>
<td>(2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• absence from work</td>
</tr>
<tr>
<td>Hoving</td>
<td>183 GP</td>
<td>12</td>
<td></td>
<td></td>
<td>• history of injury</td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• age</td>
</tr>
<tr>
<td>Boersma</td>
<td>143 GP</td>
<td>12</td>
<td>• patient expectations</td>
<td></td>
<td>• headache</td>
</tr>
<tr>
<td>(2005)</td>
<td></td>
<td></td>
<td>• negative affect</td>
<td></td>
<td>• history of injury</td>
</tr>
<tr>
<td>Mercado</td>
<td>571 gen.pop</td>
<td>6-12</td>
<td>• fear avoidance</td>
<td></td>
<td>• history of neck complaints</td>
</tr>
<tr>
<td>(2005)</td>
<td></td>
<td></td>
<td>• passive coping</td>
<td></td>
<td>• duration of complaints</td>
</tr>
</tbody>
</table>
Table 7. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Duration</th>
<th>Distress</th>
<th>Decreased Range of Motion</th>
<th>NDI Score at Baseline</th>
<th>NDI Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sterling (2005)</td>
<td>76 WAD*</td>
<td>12</td>
<td>12</td>
<td>• distress</td>
<td>• decreased range of motion</td>
<td>12</td>
</tr>
<tr>
<td>Nederhand (2004)</td>
<td>90 WAD</td>
<td>12</td>
<td></td>
<td>• fear avoidance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hendriks (2005)</td>
<td>125 WAD</td>
<td>12</td>
<td>12</td>
<td>• somatisation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GP= general practitioner, gen.pop.= general population, WAD= whiplash associated disorders, NDI= Neck Disability Index


Chapter 5

Reference List


Psychological factors in patients with sub-acute neck pain


Chapter 6

The Inter-examiner reproducibility of physical examination of the cervical spine

Published as;
Pool, Jan J.; Hoving, Jan L.; de Vet, Henrica C.; van Mameren ,Henk; Bouter, Lex M. The Inter-examiner reproducibility of physical examination of the cervical spine. Journal of Manipulative and Physiological Therapeutics 2004;27: 84-90
Abstract

Objective
To assess the inter-examiner reproducibility of physical examination of the cervical spine.

Methods
Two physiotherapists independently judged the general mobility and the inter-segmental mobility (segments C0-T2) of the neck, and the pain that was provoked. Percentage agreement and Cohen's kappa expressed agreement of dichotomous variables, limits of agreement expressed agreement of continuous variables and intra-class correlation coefficients (ICCs) expressed the reliability of continuous variables.

Results
Agreement for general mobility shows $\kappa$ between 0.05 and 0.61, and for the inter-segmental mobility it shows $\kappa$ between –0.09 and 0.63. Agreement for provoked neck pain within 1 point of an 11-point numerical rating scale (NRS) varies between 46.9% and 65.7% for general mobility and between 40.7% and 75.0% for inter-segmental mobility. The ICCs varied between 0.36 and 0.71 for general mobility and between 0.22 and 0.80 for inter-segmental mobility.

Conclusions
Despite the use of a standardised protocol to assess general mobility and inter-segmental mobility of the cervical spine it is difficult to achieve reasonable agreement and reliability between two examiners. Likewise, the patients are not able to score the same level of provoked pain in two assessments with an interval of 15 minutes.
Introduction

Neck pain is a common complaint in the general population, and its point prevalence is around 15%\textsuperscript{1}. Patients with neck pain who consult their general practitioner usually receive advice and analgesics, and approximately 43% are referred to a physical therapist or manual therapist\textsuperscript{1}. Manual assessment of the mobility of the cervical spine is made by many professionals including physical therapists, manual therapists, chiropractors and physicians. The physical examination of the cervical spine is based on the assessment of passive and/or active range of movement, including possible pain provocation during or at the end of the range of movement\textsuperscript{2}. This is assumed to provide important information with regard to the patient’s impairments. Moreover, the results of the examination and the patient’s pain response during the examination are the basis for the proposed treatment, and the results of the physical examination will be used to evaluate the treatment results\textsuperscript{2,3}. Therefore, it is important to know the reproducibility of these assessments. The reproducibility can be studied in terms of reliability and agreement. Reliability is defined as the ability to differentiate between patients and agreement is defined as the extent to which observers obtain the same measurement values in a test\textsuperscript{4,5,6,7}. Many techniques are used by physical or manual therapists to examine the cervical spine\textsuperscript{3,8} but the reproducibility of these techniques is questionable\textsuperscript{9,10,11,12,13}. Several studies have drawn different conclusions with regard to the reproducibility of manual assessment techniques\textsuperscript{9,10,11,12,13}. The majority of these studies report that better operational definitions and testing procedures are needed. The aim of this study is to investigate the inter-observer reliability and agreement of a standardised physical examination for patients with non-specific neck pain based on a protocol in which manual techniques are used to assess the general and inter-segmental mobility of the cervical spine. A new approach in this study, in comparison with former studies, is to include assessment of the inter-observer reproducibility of the patient’s pain response to the testing procedures, reported on an 11 point numerical rating scale.

Methods

During a period of 4 months (April 1999 to June 1999) 32 patients were invited to participate in the study. Patients were referred by local general practitioners in the city of Zoetermeer, in the Netherlands, to a practice providing physical and manual therapy. The patients had a
similar profile to those who participated in a recently completed randomised clinical trial on patients with neck pain\textsuperscript{14}.

Two experienced physical therapists (JJMP and LA) performed the examination of the cervical spine. The physical therapists were trained in the use of the protocol, and the movements included in the tests are part of the routine therapy they provide for patients with neck pain. They also followed the standardised protocol, which describes the performance of each of the tests in detail.

\textbf{Measurements}

Data on demographics, patient characteristics (duration, previous episodes, number of episodes), pain on a visual analogue scale and disability, using the Neck Pain Disability index\textsuperscript{15}, were collected prior to the assessment.

\textbf{General mobility of the cervical spine}

The standardised clinical assessment of the general mobility of the cervical spine consists of 6 movements;

1. full flexion and extension
2. high cervical flexion (nodding) and extension (C0-C1)
3. rotation to the right and to the left
4. lateral flexion to the right and to the left
5. combined movement A; rotation to the right and to the left, combined with extension and homo-lateral flexion (combination of the entire available movement in rotation, lateral flexion and extension)
6. combined movement B; lateral flexion to the right and to the left, combined with hetero-lateral rotation (isolating high cervical rotation).

At the end of the voluntary movement, the examiner applied a gentle passive pressure to guide the patient’s movement to the end of range to obtain a clear estimation of the range, the tissue resistance to movement and any pain response\textsuperscript{8}. The examiner classified the movements as limited or not limited, and the patient was asked to score the provoked pain at the end of each movement on an 11-point numerical rating scale (NRS), ranging from 0 (no pain at all) to 10 (extremely painful). For all movements the patient was seated on a chair with the hands on the thighs, and the back against the backrest.

\textbf{Inter-segmental mobility of the cervical spine}

The passive segmental assessment, from segment C0 to segment T2, was made with the patient in a supine position, and the examiner sitting
behind the patient. The commonly used technique to assess the segments C2 to T2 included fixation of the lower segmental level and lateral flexion to the right and to the left. Rotation was used to assess the level C1-C2, and flexion for the segment C0-C1, because these segments have different movement potentials³. The examiner classified the movement as limited or not limited and the patient was asked to score the provoked pain at the end of each movement on an 11-point numerical rating scale.

The examiners were trained in the assessment protocol, and the order of the examination was randomised according to a computer-generated random sequence table. The time-interval between the assessments was approximately 15 minutes. The examiners were blinded to each other’s results, and had no contact with each other between the assessments. In the absence of the other examiner’s results, a research assistant registered the assessment.

Assessment of reproducibility
Reproducibility was quantified in two ways: by measures of agreement, such as kappa (κ) and the Bland and Altman method⁴,⁵,⁷,¹⁶ and measures of reliability such as the intraclass correlation coefficient (ICC)⁶,⁷,¹⁷. Agreement measures assess the absolute agreement and try to quantify the measurements error. Reliability parameters assess how well persons can be distinguished from each other despite measurement errors⁷. As the mobility scores were dichotomous (limited or not limited), Cohen’s κ was used to calculate the agreement¹⁶. A κ score of 0.40 and higher was considered to be acceptable⁶,⁷,¹⁷. The provoked pain scores on the numerical rating scale were analysed as continuous variables. The ‘Bland and Altman’ method was used to assess the agreement, i.e. the extent to which examiners obtained the same measurement values in a test⁴. Using the Bland and Altman method, the inter-observer difference was calculated and plotted against the mean of the two measurements. The magnitude of the difference between the mobility scores and their distribution were visualised. The standard deviation of the difference gives an indication of the agreement of the two measurements. The 95% limits of agreement were calculated (difference ± 1.96xSDdifference), which gives an indication of the total error, i.e. bias and random error¹⁸. The percentage of agreement of the measurements was determined, allowing both 1 and 2 points difference on the numerical rating scale (0-10). For the continuous variables, the intra-class correlation coefficient (ICC), was calculated as a measure of
reliability, representing the ability to distinguish between patients\textsuperscript{6,7,17,19}. The ICC is based on a two-way random effect model ICC (2.1)\textsuperscript{6}, with the observers as a random factor (see Appendix), focusing on agreement, and ranges between 0 (no reliability) to 1 (perfect reliability). A cut-off point of ICC>0.75 was chosen a priori as an indication of acceptable reliability\textsuperscript{20}.

**Table 1:** Characteristics of patients (N=32)

<table>
<thead>
<tr>
<th>Characteristics (n=32)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (%)</td>
<td>62.5</td>
</tr>
<tr>
<td>Previous episodes of neck-complaints (%)</td>
<td>56.3</td>
</tr>
<tr>
<td>Most important complaints (%) ¶</td>
<td></td>
</tr>
<tr>
<td>- pain</td>
<td>78.1</td>
</tr>
<tr>
<td>- limitation of movement</td>
<td>40.6</td>
</tr>
<tr>
<td>- stiffness</td>
<td>28.1</td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>45.5</td>
</tr>
<tr>
<td>Mean number of episodes in the previous 5 years</td>
<td>8.6</td>
</tr>
<tr>
<td>Pain score; †</td>
<td></td>
</tr>
<tr>
<td>- mean</td>
<td>5.2</td>
</tr>
<tr>
<td>- maximum</td>
<td>7.2</td>
</tr>
<tr>
<td>- present</td>
<td>4.2</td>
</tr>
<tr>
<td>Duration of neck pain (weeks)</td>
<td>13.5</td>
</tr>
<tr>
<td>Mean NDI score ‡</td>
<td></td>
</tr>
<tr>
<td>maximal 3 complaints</td>
<td></td>
</tr>
<tr>
<td>† pain was measured on a numerical 11-point scale ranging from 0= no pain to 10=worst pain</td>
<td></td>
</tr>
<tr>
<td>‡ Neck Disability Index; disability and pain measured by 10 items ranging from 0-5 points; maximal disability 50 points</td>
<td></td>
</tr>
</tbody>
</table>

**Results**

**Patient characteristics**

During a period of 4 months (April 1999 to June 1999) 32 consecutive patients with non specific neck pain were included. The mean age of the patients was 45 years, and approximately 63% were female (Table 1). The patients had suffered from neck pain for a median duration of 13 weeks and the neck pain was recurrent in over 50% of patients. Pain was the most important complaint in 78% of the patients, limitation of movement in 40.6% of the patients; the patients were allowed to score a maximum of three complaints. The patients rated the severity of their current neck pain, on average, as 4.2 points on an 11 point numerical
rating scale. The maximum pain score in the last week was, on average, 7.2 points. The mean score for the Neck Disability Index (NDI) was 15.2 points.

Inter-examiner reproducibility of cervical mobility
The data on the general mobility are presented in Table 2. The observers scored each movement as limited or not limited. The prevalence of limited movements during the examination of the cervical spine varied from 1.6% for the high flexion end extension movements to 64.5% for the combined movement to the right. The agreement for the general cervical movement ranged from 52 - 97%, with a mean of 71%. The $\kappa$ for general cervical movement ranged from 0.05-0.61. Only rotation to the left and the combined movement A to the left showed a $\kappa$ value higher than 0.40.

Table 3 presents the data on inter-segmental mobility. The prevalence of limited segmental movements varied from 8.1% for the inter-segmental movement on the level C1-C2 on the right to 56.5% for the inter-segmental movement on the level C4-C5 on the right. The agreement for the inter-segmental movements varied from 48 - 90%, with a mean of 74%. The $\kappa$ ranged from -0.09 - 0.63. Only the levels C2-C3 and T1-T2 on the left side showed a $\kappa$ higher than 0.40.
### Table 2: Inter-examiner agreement of general mobility of the cervical spine

<table>
<thead>
<tr>
<th></th>
<th>Limited movements (prevalence)</th>
<th>Agreement %</th>
<th>Kappa 95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexion</td>
<td>21.0</td>
<td>71</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.12,0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.5</td>
</tr>
<tr>
<td>Extension</td>
<td>33.9</td>
<td>71</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.10,0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.6</td>
</tr>
<tr>
<td>High flexion C0-C1</td>
<td>1.6</td>
<td>97</td>
<td>-‡</td>
</tr>
<tr>
<td>High extension C0-C1</td>
<td>1.6</td>
<td>97</td>
<td>-‡</td>
</tr>
<tr>
<td>Rotation right</td>
<td>45.2</td>
<td>61</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.06,0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.5</td>
</tr>
<tr>
<td>Rotation left</td>
<td>38.7</td>
<td>81</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.36,0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.8</td>
</tr>
<tr>
<td>Combined movement A</td>
<td>51.6</td>
<td>55</td>
<td>0.15</td>
</tr>
<tr>
<td>right †</td>
<td></td>
<td></td>
<td>0.14,0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.4</td>
</tr>
<tr>
<td>Combined movement A</td>
<td>51.6</td>
<td>81</td>
<td>0.61</td>
</tr>
<tr>
<td>left</td>
<td></td>
<td></td>
<td>0.34,0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.8</td>
</tr>
<tr>
<td>Combined movement B</td>
<td>64.5</td>
<td>55</td>
<td>0.19</td>
</tr>
<tr>
<td>right #</td>
<td></td>
<td></td>
<td>0.01,0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.00,0.3</td>
</tr>
<tr>
<td>Combined movement B</td>
<td>53.2</td>
<td>58</td>
<td>0.20</td>
</tr>
<tr>
<td>left</td>
<td></td>
<td></td>
<td>0.11,0.5</td>
</tr>
<tr>
<td>Lateroflexion right</td>
<td>48.4</td>
<td>68</td>
<td>0.38</td>
</tr>
<tr>
<td>Lateroflexion left</td>
<td>62.9</td>
<td>52</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.24,0.3</td>
</tr>
</tbody>
</table>

† prevalence of limited movements found by examiners A and B
‡ kappa cannot be calculated because the number of limited movements was too small
† combined movement A = extension plus homo-lateral flexion plus homo-lateral rotation
# combined movement B = lateral flexion combined with a hetero-lateral rotation.
Table 3: Inter-examiner agreement of inter-segmental mobility of the cervical spine

<table>
<thead>
<tr>
<th></th>
<th>Limited movements¶ (prevalence) %</th>
<th>Agreement (%)</th>
<th>Kappa 95% CI</th>
<th>Kappa 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right</td>
<td>left</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>C0-C1</td>
<td>17.7*</td>
<td>17.7*</td>
<td>77*</td>
<td>77*</td>
</tr>
<tr>
<td>C1-C2</td>
<td>8.1</td>
<td>11.1</td>
<td>84</td>
<td>90</td>
</tr>
<tr>
<td>C2-C3</td>
<td>30.7</td>
<td>21.0</td>
<td>77</td>
<td>84</td>
</tr>
<tr>
<td>C3-C4</td>
<td>40.3</td>
<td>19.4</td>
<td>85</td>
<td>65</td>
</tr>
<tr>
<td>C4-C5</td>
<td>56.5</td>
<td>22.6</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>C5-C6</td>
<td>54.9</td>
<td>29.1</td>
<td>61</td>
<td>55</td>
</tr>
<tr>
<td>C6-C7</td>
<td>45.2</td>
<td>20.5</td>
<td>77</td>
<td>48</td>
</tr>
<tr>
<td>C7-T1</td>
<td>14.6</td>
<td>16.2</td>
<td>74</td>
<td>77</td>
</tr>
<tr>
<td>T1-T2</td>
<td>17.8</td>
<td>21.0</td>
<td>77</td>
<td>84</td>
</tr>
</tbody>
</table>

*tested movement was flexion, both joints tested at the same time, all other tested movements were lateroflexion

¶ prevalence of limited movements found by examiners A and B
Table 4: Inter-examiner agreement scores according to the Bland and Altman method and reliability (ICC) scores of provoked pain during assessment of general mobility of the cervical spine.

<table>
<thead>
<tr>
<th>Movement</th>
<th>Mean difference</th>
<th>SD of difference</th>
<th>Range of difference</th>
<th>Limits of agreement of agreement</th>
<th>Agreement ± 1 point</th>
<th>Agreement ± 2 points</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexion</td>
<td>0.13</td>
<td>2.22</td>
<td>10</td>
<td>-4.22, 4.48</td>
<td>56.3 %</td>
<td>75.1 %</td>
<td>0.63</td>
</tr>
<tr>
<td>Extension</td>
<td>0.39</td>
<td>1.96</td>
<td>10</td>
<td>-3.45, 4.23</td>
<td>62.6 %</td>
<td>81.3 %</td>
<td>0.71</td>
</tr>
<tr>
<td>Flexion C0-C1</td>
<td>0.55</td>
<td>2.67</td>
<td>14</td>
<td>-4.68, 5.78</td>
<td>46.9 %</td>
<td>68.8 %</td>
<td>0.36</td>
</tr>
<tr>
<td>Extension C0-C1</td>
<td>0.13</td>
<td>1.86</td>
<td>9</td>
<td>-3.51, 3.77</td>
<td>65.7 %</td>
<td>75.2 %</td>
<td>0.56</td>
</tr>
<tr>
<td>Rotation right</td>
<td>0.23</td>
<td>2.29</td>
<td>11</td>
<td>-4.26, 4.72</td>
<td>56.3 %</td>
<td>75.0 %</td>
<td>0.70</td>
</tr>
<tr>
<td>Rotation left</td>
<td>-0.61</td>
<td>2.33</td>
<td>10</td>
<td>-5.18, 3.96</td>
<td>50.0 %</td>
<td>62.5 %</td>
<td>0.66</td>
</tr>
<tr>
<td>Comb. mov. A Right ¶</td>
<td>0.06</td>
<td>2.43</td>
<td>9</td>
<td>-4.70, 4.82</td>
<td>50.0 %</td>
<td>65.6 %</td>
<td>0.58</td>
</tr>
<tr>
<td>Comb. mov. A Left</td>
<td>0.71</td>
<td>2.52</td>
<td>11</td>
<td>-4.23, 5.65</td>
<td>53.0 %</td>
<td>65.5 %</td>
<td>0.55</td>
</tr>
<tr>
<td>Comb. mov. B Right §</td>
<td>-1.03</td>
<td>2.78</td>
<td>10</td>
<td>-6.48, 4.42</td>
<td>46.9 %</td>
<td>69.3 %</td>
<td>0.54</td>
</tr>
<tr>
<td>Comb. mov. B Left</td>
<td>-0.45</td>
<td>2.22</td>
<td>9</td>
<td>-4.80, 3.90</td>
<td>50.0 %</td>
<td>68.8 %</td>
<td>0.65</td>
</tr>
<tr>
<td>Lateroflexion right</td>
<td>0.58</td>
<td>2.33</td>
<td>12</td>
<td>-3.99, 5.15</td>
<td>56.3 %</td>
<td>78.2 %</td>
<td>0.65</td>
</tr>
<tr>
<td>Lateroflexion left</td>
<td>-0.19</td>
<td>2.87</td>
<td>12</td>
<td>-5.86, 5.44</td>
<td>50.0 %</td>
<td>62.5 %</td>
<td>0.45</td>
</tr>
</tbody>
</table>

# Difference in provoked pain score on an 11-point rating scale between examiner A and B
‡ Intra-class correlation coefficient
¶ Combined movement A= extension plus homo-lateral flexion plus homo-lateral rotation
§ Combined movement B= lateral flexion combined with a hetero-lateral rotation
Inter-examiner reproducibility of provoked pain score

The results of inter-examiner agreement and reliability for the provoked pain scores during general cervical movements are shown in Table 4. The mean difference between observer A and B was calculated for each movement and varied between −1.03 and 0.58. The limits of agreement were broadest for lateroflexion to the left. The agreement scores within 1 point on the numerical rating scale range from 46.9 % to 65.7%, with a mean agreement of 53.7%. The mean agreement within 2 points is 70.6%. To visualise the results, the best and worse results for the inter-examiner agreement are presented in a Bland and Altman plot in Figure 1. All other plots, however, show a distribution in between these results. The ICCs for provoked pain scores ranged from 0.36 to 0.71 for the general mobility of the cervical spine, but none of the provoked pain scores reached 0.75.

The agreement of the provoked pain scores during the inter-segmental movements are shown in Table 5. The mean difference between observer A and B varied from −1.10 to 0.26, in most movements the provoked pain scores for observer B were somewhat higher. The limits of agreement were broadest for C6-C7 on the left side. The agreement score within 1 point of the numerical rating scale range from 40.6%-75.0%, with a mean agreement of 58%. The mean agreement within 2 points is 76.4%. Again, the best and the worst results of the inter-examiner agreement are plotted in Figure 2. The ICCs for inter-segmental mobility ranged from 0.22 to 0.75. An ICC higher than 0.75 was found for the levels C2-C3 and C3-C4 on the left side, and for T1-T2 on the right side.
Table 5: Inter-examiner agreement scores according to the Bland and Altman method and reliability (ICC) scores of provoked pain during assessment of inter-segmental mobility of the cervical spine

<table>
<thead>
<tr>
<th>Tested movements</th>
<th>Mean difference</th>
<th>SD difference</th>
<th>Range of difference</th>
<th>Limits of agreement ±1 points</th>
<th>Agreement ±1 points</th>
<th>Agreement ±2 points</th>
<th>ICC *</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0-C1</td>
<td>-0.57</td>
<td>1.45</td>
<td>8</td>
<td>-3.41, 2.27</td>
<td>68.7</td>
<td>87.4</td>
<td>0.73</td>
</tr>
<tr>
<td>C1-C2 R</td>
<td>-0.48</td>
<td>1.95</td>
<td>9</td>
<td>-4.30, -1.83</td>
<td>75.0</td>
<td>84.4</td>
<td>0.56</td>
</tr>
<tr>
<td>C1-C2 L</td>
<td>-0.55</td>
<td>2.83</td>
<td>15</td>
<td>-6.10, 5.00</td>
<td>65.6</td>
<td>81.1</td>
<td>0.35</td>
</tr>
<tr>
<td>C2-C3 R</td>
<td>0.03</td>
<td>2.54</td>
<td>13</td>
<td>-4.94, 5.00</td>
<td>56.4</td>
<td>81.4</td>
<td>0.50</td>
</tr>
<tr>
<td>C2-C3 L</td>
<td>-0.13</td>
<td>1.87</td>
<td>9</td>
<td>-3.80, 3.53</td>
<td>63.2</td>
<td>82.0</td>
<td>0.78</td>
</tr>
<tr>
<td>C3-C4 R</td>
<td>-0.52</td>
<td>2.32</td>
<td>11</td>
<td>-5.07, 4.03</td>
<td>62.5</td>
<td>75.1</td>
<td>0.62</td>
</tr>
<tr>
<td>C3-C4 L</td>
<td>-0.16</td>
<td>2.05</td>
<td>9</td>
<td>-4.18, 3.85</td>
<td>59.4</td>
<td>78.2</td>
<td>0.75</td>
</tr>
<tr>
<td>C4-C5 R</td>
<td>-0.83</td>
<td>2.19</td>
<td>11</td>
<td>-5.12, 3.46</td>
<td>53.1</td>
<td>78.1</td>
<td>0.62</td>
</tr>
<tr>
<td>C4-C5 L</td>
<td>0.26</td>
<td>2.64</td>
<td>13</td>
<td>-4.91, 5.43</td>
<td>40.6</td>
<td>68.7</td>
<td>0.55</td>
</tr>
<tr>
<td>C5-C6 R</td>
<td>-0.87</td>
<td>2.16</td>
<td>10</td>
<td>-5.10, 3.36</td>
<td>62.6</td>
<td>75.1</td>
<td>0.66</td>
</tr>
<tr>
<td>C5-C6 L</td>
<td>-0.16</td>
<td>2.16</td>
<td>9</td>
<td>-4.30, 4.07</td>
<td>56.3</td>
<td>71.9</td>
<td>0.65</td>
</tr>
<tr>
<td>C6-C7 R</td>
<td>-0.29</td>
<td>2.53</td>
<td>12</td>
<td>-5.25, 4.67</td>
<td>53.2</td>
<td>72.0</td>
<td>0.59</td>
</tr>
<tr>
<td>C6-C7 L</td>
<td>-0.35</td>
<td>3.46</td>
<td>16</td>
<td>-7.13, 6.43</td>
<td>47.0</td>
<td>59.5</td>
<td>0.22</td>
</tr>
<tr>
<td>C7-T1 R</td>
<td>-1.03</td>
<td>2.59</td>
<td>16</td>
<td>-6.11, 4.05</td>
<td>59.5</td>
<td>75.1</td>
<td>0.45</td>
</tr>
<tr>
<td>C7-T1 L</td>
<td>-0.94</td>
<td>2.82</td>
<td>16</td>
<td>-6.47, 4.59</td>
<td>56.3</td>
<td>75.1</td>
<td>0.34</td>
</tr>
<tr>
<td>T1-T2 R</td>
<td>-1.03</td>
<td>1.43</td>
<td>6</td>
<td>-3.83, 1.77</td>
<td>56.3</td>
<td>81.3</td>
<td>0.80</td>
</tr>
<tr>
<td>T1-T2 L</td>
<td>-1.10</td>
<td>2.28</td>
<td>10</td>
<td>-5.57, 3.37</td>
<td>56.2</td>
<td>71.8</td>
<td>0.54</td>
</tr>
</tbody>
</table>

* Difference in provoked pain score on an 11-point rating scale between examiners A and B

* Intra-class correlation coefficient
**Figure 1.** Agreement of provoked pain scores on a numerical rating scale during assessment of general mobility between examiners. Dotted lines; limits of agreement Straight lines; mean difference between examiners
Figure 2. Agreement of provoked pain scores on a numerical rating scale during assessment of inter-segmental mobility between examiners
Dotted lines; limits of agreement
Straight lines; mean difference between observers
Discussion

Despite considerable training and the use of a standardised protocol, the results of this study show that the reproducibility of cervical mobility and pain provoked during mobility assessments was highly variable and overall unacceptable. The assessment of inter-segmental mobility showed a slightly better agreement, followed by the agreement for general mobility. However the κ values were disappointing. The explanation for the differences in agreement and Cohen’s κ (the agreement corrected for chance) could be the unequal prevalence of positive or negative findings\(^\text{10,17}\). Fjellner\(^\text{10}\), for example, only calculated the κ if the mean prevalence of positive findings for two examiners was between 10% and 90%. In this study the distribution of negative findings (ie, movement not limited), was more than 90% for high cervical flexion and extension.

For the general mobility of the cervical spine the standard deviation of the difference between provoked pain scores varied between 2 and 3 points. The limits of agreement, which give an indication of the measurement error, vary from 4 to 6 points. These limits of agreement are much too wide to label it as an acceptable agreement. The pain scores for the inter-segmental movements show similar limits of agreement as for the general movements, with the exception of a systematic difference between examiner A and examiner B. (the scores of examiner B being systematically higher).

An ICC score of 0.75 was defined as an acceptable level of reliability, and in this study only 3 of the 29 measurements met that criteria.

Studies focussing on the reproducibility of methods to assess the cervical spine are rare. Most studies that examined the mobility of the cervical spine have reported that it is difficult to achieve a reasonable score for agreement and reliability. Jull et al\(^\text{21}\) studied inter-examiner agreement in detecting cervical joint dysfunction, and reported very high κ values. Viikari\(^\text{13}\) studied a total assessment of the neck with a conventional neurological examination, palpation of the neck and shoulder region and clinical tests consisting of 34 items, and concluded that only 4 items showed good agreement. Strender et al\(^\text{12}\) evaluated the inter-examiner agreement of 10 clinical tests, and found that only 2 tests had an acceptable agreement. Fjellner et al\(^\text{10}\) evaluated a number of clinical tests for the assessment of passive and inter-segmental movement and found an acceptable agreement in tests of passive general movement, but in a very few of the tests for passive inter-segmental movement. Smedmark et al\(^\text{11}\) studied 4 tests performed on
61 patients and found a relatively high percentage of agreement, but fair to moderate κ.

The studies carried out by Strender et al and Fjellner et al focussed on normal healthy subjects so the results may therefore not apply to patients with neck pain. The majority of the above mentioned studies reported that better operational definitions and testing procedures were needed.

The assessment with the most reliable score reported in the literature is the foramen compression test (κ=0.43)\textsuperscript{12}, but the combined movement A in the present study which involves a similar movement, has a low kappa and agreement on the right side.

Because different techniques are used in daily practice, it is difficult to make comparisons between studies. Furthermore, only a few studies have focussed on patients with neck pain, and the agreement and reliability of provoked pain scores have not been studied before. Although the pain score during an assessment is rather subjective\textsuperscript{2}, a classification of a dysfunction is made on the basis of the parameters of pain or restricted or limited movement\textsuperscript{2}. For most patients it was very hard to report the same pain score. A higher pain score might have been expected for the second assessment, but no systematic difference was found in the scores. The variation is probably due to the variation between the examiners\textsuperscript{9,20}, for example because of differences in palpation and movements of the same level of the cervical spine and/or to the force used for over-pressure. Earlier studies suggested to improve reproducibility in daily practice by standardisation of the examination protocol. However even with extensively trained physiotherapist we found unsatisfying results.

Assessments of general or segmental mobility in daily practice are poorly reproducible and therefore to diagnose only on the outcome of such a assessment is not recommendable. More research is needed in the search for reliable instruments and techniques in daily practice. For research purposes it is always possible to increase the sample size, which is a strategy to cope with measurements with a large amount of random error.

**Conclusion**

Despite the use of a standardised protocol to assess general mobility and inter-segmental mobility of the cervical spine it is difficult to achieve reasonable agreement and reliability between two examiners. Likewise, the patients are not able to score the same level of provoked pain in two assessments with an interval of 15 minutes.
References


Reproducibility of cervical range of motion in patients with neck pain

Published as;
BMC. Musculoskelet. Disord. 2005 dec 13;6:59
Abstract

Objective: To assess the intra-rater and inter-rater reproducibility of the measurement of active Range of Motion (ROM) in patients with neck pain using the Cybex Electronic Digital Inclinometer-320 (EDI-320).

Design: In an outpatient clinic in a primary care setting 32 patients with at least 2 weeks of pain and/or stiffness in the neck were randomly assessed, in a test-retest design with blinded raters using a standardized measurement protocol.

Main outcome measure: Cervical flexion-extension, lateral flexion and rotation was assessed.

Results: Reliability expressed by the Intraclass Correlation Coefficient (ICC) was 0.93 (lateral flexion) or higher for intra-rater reliability and 0.89 (lateral flexion) or higher for inter-rater reliability. The 95% limits of agreement for intra-rater agreement, expressing the range of the differences between two ratings were -2.5 ± 11.1° for flexion-extension, -0.1 ± 10.4° for lateral flexion and -5.9 ± 13.5° for rotation. For inter-rater agreement the limits of agreement were 3.3 ± 17.0° for flexion-extension, 0.5 ± 17.0° for lateral flexion and -1.3 ± 24.6° for rotation.

Conclusions: In general, the intra-rater reproducibility and the inter-rater reproducibility were good. We recommend to compare the reproducibility and clinical applicability of the EDI-320 inclinometer is compared with other cervical ROM measures in symptomatic patients.
Introduction

Neck pain is a common musculoskeletal disorder. The point prevalence for neck pain in the general population of the Netherlands varies between 9% and 22%\(^5\);\(^19\), and approximately one-third of all adults will experience neck pain during the course of 1 year\(^9\). Patients usually receive conservative treatment such as physical therapy or continued care by a General Practitioner (GP)\(^4\). A physical evaluation is often used for both the diagnosis and the evaluation of treatment success in patients with neck pain\(^14\). One aspect for the physical assessment of the cervical spine is the evaluation of active Range Of Motion (ROM). Active cervical ROM is difficult to measure because of compensatory movements, and it is influenced by aging and systemic disorders\(^24\). Several non-invasive methods for assessing the ROM have been available, such as visual estimation, two-arm goniometry, inclinometry, compass technology, video technology, electromagnetic technology and potentiometry. For the majority of these instruments the intra- and inter-rater reproducibility has not been tested adequately. Radiography has been proven to be of questionable reproducibility\(^7\);\(^24\).

In an extensive critical appraisal of reliability studies on cervical ROM measures Jordan evaluated 21 papers for methodological rigor\(^15\). Commonly identified flaws in these reliability studies were low sample size, unclear selection criteria, the use of only healthy individuals, use of inadequate reliability statistics, the absence of a protocol, and questionable applicability in clinical practice. In our experience the Cybex Electronic Digital Inclinometer-320 (EDI-320) is a practical tool for the objective measurement of active ROM\(^8\). One of the clinical advantages of the EDI-320 is that it does not have to be fitted on the patient and it is portable.

Previous studies using the EDI-320 have investigated the intra-rater and inter-rater reproducibility only in healthy subjects\(^16\);\(^23\). It is unknown whether these reproducibility results are applicable to patients with pain or stiffness in the neck. Consequently, the aim of our study is to determine the intra-rater and inter-rater reproducibility in patients with non-specific neck pain. We also assess whether the reproducibility can be improved when two ratings per rater are used instead of one rating. Furthermore, we evaluate whether the inter-rater reproducibility is affected by the severity of pain. Results of reproducibility studies can be used for many purposes. One application is the determination of changes that can be detected beyond
Chapter 7

measurement error: the smallest detectable difference (SDD). In the present study we assess SDD for an individual patient.

Methods

Patient characteristics
Consecutive patients with neck pain, referred by local general practitioners for physical therapy in Zoetermeer, the Netherlands, were invited to participate. The selection criteria were: age between 18 and 70 years, pain and/or stiffness in the neck for at least 2 weeks, and written informed consent. Patients were excluded if they had received surgery in the cervical region or had evidence of specific pathology, such as malignancy, neurological disease, fracture, herniated disc or systemic rheumatic disease. Data on demographics (e.g. age and gender), clinical factors (duration, concomitant complaints), neck pain on a numerical 0-10 point scale ranging from 0 (no pain) to 10 (maximal pain), and disability assessed with the Neck Disability Index (NDI) were collected by an independent research assistant prior to the actual active ROM measurements.

Rater characteristics
The raters were two physical therapists with 3 months experience using the EDI-320 inclinometer (Lumex, Inc., Ronkonkoma, New York) and both performed weekly cervical ROM assessments in another study. The measurement procedures were practiced on 5 healthy volunteers prior to the start of the present study.

Measurement protocol
For the measurements of cervical flexion-extension and lateral flexion the patient was seated upright in a high chair, with the hands resting on the upper thigh. For the measurement of cervical flexion-extension, the position of 0 degrees was in maximal cervical flexion (“chin to chest”), followed by maximal cervical extension. Likewise, the measurements of lateral flexion were initiated with the position of 0 degrees in maximal lateral flexion to the left (“ear to left shoulder”), followed by maximal lateral flexion to the right. Because active ROM using the EDI-320 inclinometer can only be measured against gravity, the ratings of cervical rotation were performed with the patient in a supine position. The position of 0 degrees was in maximal left rotation, followed by maximal right rotation. During rotation the head slide over a cushioned
treatment table and the patient was not allowed to make any compensatory lateral flexion with the head. See Figure 1.

Figure 1. Rotation to the left

We chose for full cycle ROM (for example: from left to right rotation) as the neutral head position is difficult to perform in half-cycle ROM (for example: from the neutral to left rotation) assessments in the cervical spine\(^7\). The reference point for the EDI was on the forehead for both flexion-extension and rotation, and right above the ear for lateral flexion. Throughout the motion, the physiotherapist maintained contact with the EDI and the reference point on the head.

The subjects were instructed to perform the movement and then to practice twice before performing the actual movement. The patient was instructed only to move the head, and to avoid compensatory movements in the thoracic or lumbar region. The patient was gently guided through the whole range of motion, and manual contact was applied by the rater. The patient was encouraged to perform a maximal movement until the end of the active ROM was reached, or until the pain prevented the patient from going any further.
Procedure reproducibility study
Active ROM of the cervical spine was assessed twice in three planes in the following order: maximal flexion to maximal extension (2x), maximal lateral flexion from left to right (2x), and maximal rotation from left to right (2x). The time interval between measurement between the first and second ratings of a single rater was 5 minutes and the interval between raters was 10 minutes. The order of the raters was randomized using a computer generated random sequence table. At all times only one rater was present in the examination room, together with the research assistant. The research assistant recorded the number of degrees, which were electronically displayed on the EDI-320. In order to keep the raters blind for the outcome of measurement, the read out on the electronic display of the EDI-320 was concealed for both raters and patients. Thus, the raters were unaware of the previous measurements by the other rater.

Data Analysis
We used two different measures which are increasingly used in reproducibility studies: one measure to assess reliability and one measure to assess agreement\textsuperscript{11,20}. Figure 1 shows an overview of the intra-rater and inter-rater comparisons we made.

Agreement parameters
Parameters of agreement measure the ability to achieve the same value in two measurements, and gives an indication of the size of the measurement errors. We assess the 95% limits of agreement (LoA) according to Bland and Altman as a measure of agreement\textsuperscript{3}. The mean difference between the scores of both raters was calculated, representing the systematic differences (bias) between the measurements. The standard deviation (SD) of this difference represented the extent to which the rater(s) recorded the same mean value in each plane. Then the 95% limits of agreement (LoA) were calculated (mean of the difference ± 1.96*SD), indicating the ‘total error’, systematic and random error together\textsuperscript{1}.

As no clear criteria exist for acceptable value of intra-rater and inter-rater agreements for active ROM outcome measures, we defined, a priori, that a difference in measurement between the raters of 10% of the total range of measurement values would be acceptable.

The Bland and Altman method can be visualized by plotting the differences between the first and the second ratings against the
Reproducibility of cervical range of motion

corresponding mean of the first and the second rating. This visual representation of agreement illustrates the magnitude and range of the differences, bias or outliers, and the relation between the magnitude of the differences and the magnitude of the mean values.

Based on the agreement results of rater A the smallest detectable difference (SDD) for an individual level was calculated for each movement by multiplying the SD of the differences by 1.96: 1.96* SD change. The SDD represents the change that can be detected by the EDI-320 beyond measurement error.

Reliability parameters

Reliability parameters reflect the extent to which a measurement instrument can differentiate between patients. If persons differ a lot, it is easier to distinguish them from each other, despite some measurement errors. In that case the measurement errors are related to the differences between the persons.

As a parameter of reliability the Intraclass Correlation Coefficient (ICC) was used (Figure 1). We used ICCs which took systematic differences in the measurements into account. These ICCs are defined as the ratio of the variance among patients (patient variability) over the total variance (among patients, among raters plus the error variance), which ranges between 0 (no reliability) and 1 (perfect reliability). The cut-off point of ICC>0.75 was chosen a priori as an indication of acceptable reliability. We used SPSS 9.0 statistical software (SPSS Inc., Chicago, Illinois) to calculate the ICCs. In case the unit of analyses was the mean of two ratings by one rater, variances in which the raters were involved were divided by a factor 2.

Figure 2 shows an overview of intra- and inter-rater comparisons:

**Figure 2.** Flow diagram measurements and assessment of intra- and inter-rater reproducibility
Results

Patient characteristics

During a period of 4 months (April 1999 - June 1999) 32 patients with neck pain were recruited. The mean age of patients included in this study was 45 years, and approximately 63% were female (Table 1). Patients had suffered from neck pain for a median duration of 13 weeks and in more than half of the patients the neck pain was recurrent. Patients rated the severity of their current neck pain, on average, as 4.2 on a numeric 11-point scale. The mean score for the NDI was 15.2 points (maximal disability: 50 points).

Table 1. Characteristics of patients

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency* (n = 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, sd)</td>
<td>45.5 (9.2)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (62.5)</td>
</tr>
<tr>
<td>Headache</td>
<td>2 (6.3)</td>
</tr>
<tr>
<td>Trauma reported as cause</td>
<td>8 (25.0)</td>
</tr>
<tr>
<td>Reported stiffness of the neck</td>
<td>9 (28.1)</td>
</tr>
<tr>
<td>Previous neck pain episodes</td>
<td>18 (56.3)</td>
</tr>
<tr>
<td>Current pain 0-10 (mean, sd) †</td>
<td>4.2 (2.3)</td>
</tr>
<tr>
<td>NDI score (mean, sd) ‡</td>
<td>15.2 (8.3)</td>
</tr>
<tr>
<td>Duration neck pain (median; IQR§)</td>
<td>13.5 (8.0, 25.5)</td>
</tr>
</tbody>
</table>

* Number of patients and % between brackets unless stated otherwise
† Current pain was measured on a numerical 11-point scale ranging from 0=no pain to 10=worst pain.
‡ Neck Disability Index: disability and pain measured by 10 items ranging from 0-5 points. Maximal disability 50 points
§ IQR= Inter-Quartile Range

Intra-rater agreement

The intra-rater agreement and reliability results are shown in Table 2 and Figure 3. Small but statistically significant systematic differences are seen for rotation for which the second rating of active ROM is higher, both for rater A and B (mean difference rater A: -5.9, 95% CI -8.4 to -3.4; rater B: -2.7, 95% CI -5.3 to -0.03) and for flexion-extension (rater A: -2.5, 95% CI -4.5 to -0.5). The limits of agreement were broadest for rotation. The standard deviation of the difference, representing the extent to which rater A achieved the same mean scores for the first and second rating, ranged between 5.3º (lateral flexion),
5.7° (flexion-extension) and 6.9° (rotation). From these, the limits of agreement were calculated. For any new patient it is expected with an approximate 95% probability that the difference between the two ratings of rater A should lie within the limits of agreement; which were – 2.5±11.1° for flexion-extension, -0.1±10.4° for lateral flexion and – 5.9±13.5° for rotation. Figure 3 shows that the magnitude of the difference is not associated with their mean value, indicating that the mean difference and the standard deviation of the differences are adequate summary statistics of agreement.
<table>
<thead>
<tr>
<th>Tested movements</th>
<th>1st rating mean</th>
<th>SD</th>
<th>2nd rating mean</th>
<th>SD</th>
<th>1st-2nd rating mean</th>
<th>SD</th>
<th>LoA ± 95% CI</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rater A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>101.3</td>
<td>21.6</td>
<td>103.8</td>
<td>23.0</td>
<td>-2.5</td>
<td>5.7</td>
<td>-2.5±11.1</td>
<td>0.96</td>
<td>0.93, 0.98</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>72.6</td>
<td>13.8</td>
<td>72.7</td>
<td>14.4</td>
<td>-0.1</td>
<td>5.3</td>
<td>-0.1±10.4</td>
<td>0.93</td>
<td>0.86, 0.97</td>
</tr>
<tr>
<td>Rotation</td>
<td>135.4</td>
<td>30.4</td>
<td>141.3</td>
<td>29.7</td>
<td>-5.9</td>
<td>6.9</td>
<td>-5.9±13.5</td>
<td>0.96</td>
<td>0.91, 0.98</td>
</tr>
<tr>
<td><strong>Rater B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>99.8</td>
<td>21.0</td>
<td>98.8</td>
<td>22.7</td>
<td>1.0</td>
<td>5.6</td>
<td>1.0±11.1</td>
<td>0.97</td>
<td>0.93, 0.98</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>71.8</td>
<td>12.3</td>
<td>72.4</td>
<td>14.1</td>
<td>-0.6</td>
<td>5.0</td>
<td>-0.6±9.8</td>
<td>0.93</td>
<td>0.86, 0.96</td>
</tr>
<tr>
<td>Rotation</td>
<td>138.3</td>
<td>27.0</td>
<td>141.0</td>
<td>28.9</td>
<td>-2.7</td>
<td>7.4</td>
<td>-2.7±14.4</td>
<td>0.96</td>
<td>0.92, 0.98</td>
</tr>
</tbody>
</table>

*LoA = 95% Limits of Agreement, ICC = Intra-class Correlation Coefficient, 95% CI = 95% Confidence Interval

**Table 2.** Intra-rater reproducibility analyses*
### Table 3. Inter-rater reproducibility analyses *

<table>
<thead>
<tr>
<th>Tested movements</th>
<th>Rater A</th>
<th>Rater B</th>
<th>Rater A-B</th>
<th>LoA</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SD</td>
<td>mean</td>
<td>SD</td>
<td>mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Mean of two ratings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>102.6</td>
<td>22.2</td>
<td>99.3</td>
<td>21.7</td>
<td>3.3</td>
<td>8.7</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>72.6</td>
<td>13.9</td>
<td>72.1</td>
<td>13.0</td>
<td>0.5</td>
<td>8.7</td>
</tr>
<tr>
<td>Rotation</td>
<td>138.4</td>
<td>29.9</td>
<td>139.7</td>
<td>27.7</td>
<td>-1.3</td>
<td>12.5</td>
</tr>
<tr>
<td><strong>One rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>101.3</td>
<td>21.6</td>
<td>99.8</td>
<td>21.0</td>
<td>1.6</td>
<td>10.1</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>72.6</td>
<td>13.8</td>
<td>71.8</td>
<td>12.3</td>
<td>0.8</td>
<td>8.9</td>
</tr>
<tr>
<td>Rotation</td>
<td>135.4</td>
<td>30.4</td>
<td>138.3</td>
<td>27.0</td>
<td>-2.9</td>
<td>13.7</td>
</tr>
<tr>
<td><strong>Low pain intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>111.0</td>
<td>17.5</td>
<td>106.8</td>
<td>16.1</td>
<td>4.3</td>
<td>8.2</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>77.0</td>
<td>11.3</td>
<td>76.0</td>
<td>11.6</td>
<td>1.1</td>
<td>8.9</td>
</tr>
<tr>
<td>Rotation</td>
<td>147.7</td>
<td>20.7</td>
<td>148.0</td>
<td>17.4</td>
<td>-0.3</td>
<td>12.8</td>
</tr>
<tr>
<td><strong>High pain intensity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexion-extension</td>
<td>81.0</td>
<td>18.1</td>
<td>80.2</td>
<td>23.1</td>
<td>0.8</td>
<td>9.8</td>
</tr>
<tr>
<td>Lateral-flexion</td>
<td>61.4</td>
<td>14.1</td>
<td>62.3</td>
<td>11.5</td>
<td>-0.8</td>
<td>8.4</td>
</tr>
<tr>
<td>Rotation</td>
<td>114.6</td>
<td>37.3</td>
<td>118.4</td>
<td>37.8</td>
<td>-3.8</td>
<td>12.1</td>
</tr>
</tbody>
</table>

* LoA = 95% Limits of Agreement, ICC = Intra-class Correlation Coefficient, 95% CI = 95% Confidence Interval
Chapter 7

Cervical rotation

Mean rotation 1st and 2nd measurement

Difference 1st and 2nd measurement

<table>
<thead>
<tr>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>0</th>
<th>-10</th>
<th>-20</th>
<th>-30</th>
<th>-40</th>
</tr>
</thead>
</table>

Cervical lateral flexion

Mean lateral flexion 1st and 2nd measurement

Difference 1st and 2nd measurement

<table>
<thead>
<tr>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>0</th>
<th>-10</th>
<th>-20</th>
<th>-30</th>
<th>-40</th>
</tr>
</thead>
</table>

Presented is the line of the mean difference (in the middle) and the limits of agreement (the two dotted outer lines corresponding with the +2 SD’s or - 2 SD’s of the mean difference between the 1st and 2nd rating by rater A)

Figure 3. Plotted differences within rater A against the mean value of the first and second measurement for each patient for cervical flexion-extension, lateral flexion and rotation

Inter-rater agreement

Regarding the inter-rater agreement, only minor systematic differences were observed between rater A and rater B for lateral flexion and rotation (Table 3). For flexion-extension there is a small but statistically significant systematic difference (mean 3.3; 95% CI 0.2 to 6.4). The
Reproducibility of cervical range of motion

standard deviation of the difference ranged between 8.7° (flexion-extension and lateral flexion) and 12.5° (rotation). The limits of agreement were 3.3±17.0° for flexion-extension, 0.5±17.0° for lateral flexion and -1.3±24.6° for rotation.

Intra-rater and inter-rater reliability
The intra-rater reliability was high with ICCs ranging from 0.93 (lateral flexion rater A and B) to 0.97 (flexion-extension rater B). Likewise the inter-rater reliability was also good with ICCs of 0.89 or higher for all three planes.

One rating versus two ratings per rater
Table 3 shows that when only one rating per rater was used instead of two, the limits of agreement were slightly wider and the ICC were slightly lower.

The influence of pain on the inter-rater agreement and reliability
In addition, we compared patients with a high pain score (7 points or higher on a 0-10 point scale, n=9) to patients with a low or moderate pain score (6 points or lower on a 0-10 point scale, n=23). Patients with high pain intensity had lower active ROM values compared to patients with a low pain intensity (p≤0.05). Although the standard deviations of the individual raters were higher in the high pain intensity group, the standard deviations of the mean differences were similar and consequently, the 95% limits of agreement did not differ much (Table 3). Therefore also the limits of agreement are similar. The ICC values in the high pain intensity group are slightly higher compared with those in the low pain intensity group.

The smallest detectable difference
The mean active ROM values (mean of 4 ratings by 2 raters) were 100.9 degrees for flexion-extension, 72.4 degrees for lateral flexion and 139.0 degrees for rotation. The acceptable differences to be detected, defined as 10% of the used range of the scale, were therefore 10.1 for flexion-extension, 7.2 degrees for lateral flexion and 13.9 degrees for rotation. Based on the intra-rater agreement results (rater A), the SDD for an individual was 11.1 degrees for flexion-extension, 10.4 degrees for lateral flexion and 13.5 degrees for rotation. This means that only changes in cervical range of motion larger than these values can be detected beyond measurement error when a single physiotherapist performs both measurements. If the measurements on which the change in cervical range of motion is based are performed by two
different raters than the SDDs were 17.0, 17.0 and 24.6 for flexion-extension, for lateral flexion and for rotation, respectively.

**Discussion**

The first aim of this study was to investigate the intra- and inter-rater reproducibility of the assessment of range of motion in three planes for patients with neck pain, using the Cybex EDI-320 inclinometer. For intra-rater reproducibility we compared the first rating with the second rating of each rater and for the inter-rater reproducibility we compared rater A with rater B. Some systematic differences were observed, however these were small considering the overall active ROM in each plane for both the intra-rater and inter-rater agreement. Overall, we found good intra- and inter-rater reliability statistics (ICCs of 0.86 or higher). As expected both agreement and reliability were slightly higher for the intra-rater comparisons than for the inter-rater comparisons. High reliability does not necessarily mean that the raters agree in an absolute sense on the active ROM (agreement)\(^{11,20}\). For this reason we included both parameters of agreement and reliability in the present study.

The SDD, based on intra-rater agreement, for flexion-extension (11.1°) and rotation (13.5°) was almost equal to the cut-off values for our predefined criteria for an acceptable clinical difference (10.1° and 13.9°, respectively). However, for lateral flexion (10.4°) an acceptable clinical difference may be somewhat more difficult to detect as the SDD was higher than our predefined acceptable difference of 10% (7.2°). Also measurements performed by different raters are insufficiently reproducible to detect the predefined difference of 10% of the used range of the measurement scale. However, this holds for SDDs calculated on the individual level. In research, when groups of patients are used the EDI-320 is sufficiently reproducible for all measurements of range of motion, because SDD values should be divided by \(\sqrt{N}\) to obtain SDD for group level, with a group size of N.

To minimize any random error, the inter-rater statistics were based on the mean of two ratings as outlined in our protocol. We investigated whether just one rating per rater instead of two would yield acceptable reproducibility statistics (second aim). Although a duplicate rating did not improve the reproducibility much, the 2nd rating with the EDI-320 can be done easily. Similarly, we evaluated whether reproducibility was affected by the severity of pain. Patients with high pain intensity had on
average, less ROM compared to patients in the low pain intensity group (p≤0.05). However, reliability and agreement were acceptable in both the group with low and high pain intensity.

We hypothesized that pain and limitation of movement could either increase or decrease during the course of a series of movements and thus pose sources of systematic variation to the assessment of reproducibility. By comparing the first and second consecutive pair of ratings (independent of the rater), a statistically significant small, but not clinically relevant, difference was observed for flexion-extension (3.4 degrees difference: 95% CI 0.2 to 6.5). We therefore conclude that the effect of repeated movements on cervical ROM was minimal.

In the present study we looked at the intra-rater reproducibility by comparing two consecutive ratings with a minimal time interval and inter-rater reproducibility with an interval of approximately 10 minutes. The main reason for the choice of the time interval of 10 minutes was a practical one: we could measure a patient in one single visit. Our assumption was that within 10 minutes the patients will be stable on pain perception and range of motion. Had we chosen a larger time interval our results might have been different, however. Ideally, true intra-rater variability is evaluated for a disorder stable within the time frame evaluated. However, we consider a large time interval not desirable for the assessment of measurement variation because of the biological variation within subjects over time.

More than half of all studies on the reproducibility of cervical ROM have inappropriately used T-tests or repeated measures ANOVA, which are not considered true reliability statistics. The ICC is used in only a few studies. ICC values are known to be dependent on the variation in the study population. As can be seen from the visual representation of agreement (Figure 2), the active ROM values for lateral flexion are somewhat more clustered together (a smaller range) than the other two planes. The more homogeneous values might give some explanation for the somewhat lower ICCs for lateral flexion, and the wider range of values result in higher ICCs for rotation. Likewise, the larger variation in active ROM values in the high pain intensity group might also explain the higher ICCs compared to the low pain intensity group.

Studies that measure ROM for patients with neck disorders are scarce. A systematic review identified that only 6 studies assessed reliability in patients with cervical disorders and of these only 2 studies had more than 30 subjects. Two studies reported on the reproducibility of the...
EDI-320 for cervical ROM in healthy subjects\textsuperscript{16;23}. The first one reported acceptable agreement results and found that more than 90\% of the successive ratings for cervical flexion and lateral flexion by two raters were within a range between 0-10 degrees\textsuperscript{15}. The other study only investigated flexion and extension, and reported moderate to high intra-rater reliability (flexion ICC 0.77, extension 0.79-83) and somewhat lower inter-rater reliability (flexion ICC 0.66-0.73; extension ICC 0.66-0.80)\textsuperscript{23}. The authors of this study report that the reliability could be improved by using a standardized protocol. Comparison of ICC values between different studies is hampered by the dependency of ICC values on the variability of range of motion values of the population under study\textsuperscript{11}. De Winter et al showed that for measurements of range of motion in 155 patients with shoulder complaints, the ICC were high for the affected shoulder (ICC = 0.83) and low for the non-affected shoulder (ICC=0.28). This difference was completely due to variability of range of motion found for the affected shoulder, which was large and the non-affected shoulder, which was low.

The CROM device is the most frequently reported measure for cervical ROM and variable ICC values have been reported, both alone or when compared to other ROM instruments\textsuperscript{6;15;26}. One study on patients with cervical spine disorders reported inter-rater ICCs for active ROM greater than 0.80 with the Cervical Range of Motion Device (CROM device) compared to ICCs lower than 0.80 for visual estimation and a universal goniometer\textsuperscript{26}. Considering the results of this study it would be interesting to directly compare the CROM device with the EDI-320 inclinometer in a future study.

Our population consisted of patients with non-specific neck pain, readers can compare the patient profile presented in this article with their own patients. The measurement procedure is quick and simple, which we hope will facilitate replication of our reproducibility design in other clinical settings.

**Conclusion**

In general, the intra-rater reproducibility and the inter-rater reproducibility were acceptable, despite slight variations. We recommend that the reproducibility and clinical applicability of the EDI-320 inclinometer is compared with other cervical ROM measures in a symptomatic patient population.
Reference List


The applicability of the Tampa Scale of Kinesiophobia for patients with sub-acute neck Pain: a qualitative study

Chapter 8

Abstract

The purpose of this study was to qualitatively evaluate patients' understanding and interpretation of the wording used in test items of the Tampa Scale of Kinesiophobia (TSK). The TSK was developed to measure fear of movement in patients suffering from low back pain. The TSK is being increasingly used for other pain conditions. Patients with sub-acute neck pain experience problems while completing this questionnaire. The aim of this study was to elicit these problems.

The study was conducted within the framework of a randomised controlled trial. The Three-Step Test Interview (TSTI) was used to collect data on the thoughts or considerations of respondents while completing the TSK. In the analysis, each transcribed interview was divided into three segments. The thoughts and considerations were then analysed and categorised per item.

During the TSTI two problems were identified. One concerned the meaning of specific words used, like “dangerous” and “injury”. The other problem was that several implicit assumptions within some items make it difficult for respondents to answer these items.

It was concluded that in the development and validation of questionnaires like the TSK, not only quantitative psychometric properties are important, but also qualitative research has an important contribution to enhance applicability.
**Introduction**

Patients suffering pain often report fears that they are afraid that movement will exacerbate their symptoms. Fear of pain, and fear of movement or re-injuring tissues are important factors in the development and maintenance of pain. Fear might contribute to how a patient moves, behaves and experiences his or her pain. The transition from acute to chronic pain is believed to be influenced by fear of movement, and among many other factors such as cognitive, behavioural and social factors. The fear-avoidance model suggests possible pathways by which pain patients become enmeshed in a downward spiral of increasing avoidance, disability and pain. Although most research on the fear-avoidance model has been carried out in chronic low back pain patients, this model also yields an interesting perspective on the development of chronic neck pain.

The Tampa Scale for Kinesiophobia (TSK) was developed and translated into Dutch by Vlaeyen et al (see Appendix for the English version). The TSK is a psychological self-completion questionnaire consisting of 17 items, and, each item is scored on a 4-point Likert scale with response options ranging from 'strongly agree' to 'strongly disagree'. Sum-scores range from 17 to 68, with higher scores indicating more fear of movement and/or (re-)injury. Four items are phrased in reversed key (items 4, 8, 12 and 16). Several studies have found evidence for the validity, i.e. predictive validity and construct validity, and the reliability, i.e. internal consistency and test-retest reliability in chronic and acute low back pain patients, of the Dutch version of the TSK.

To be valid and applicable, the items in a questionnaire should measure what they are supposed to measure and therefore be fully understood. Nevertheless, in practice not all items meet this prerequisite. A possible explanation could be that respondents do not understand the items in a questionnaire, or misinterpret the items or the response options. In other words, like most social research tools, they are open to various different interpretations. Quantitative psychometric analyses, such as factor analysis, internal consistency and construct validity, will highlight some problems with the structure or formulation of items of a questionnaire but they are not very sensitive with regard to problems concerning the way in which patients interpret items, or their intended meaning when they select a response. The basic assumption is that patients are able to understand the meaning of the items, that items are
understood in the same way by all patients, and that all patients are willing and able to respond to all items.

During a randomised clinical trial on neck pain patients, some of the TSK items appeared to cause problems. Some patients had to think for a long time about specific items, they did not understand some items or specific words in certain items, and they asked the research assistants for additional information before they could respond to some of the items.

The objective of the present qualitative study was to find out what problems patients with sub-acute neck pain encounter when completing the TSK. Therefore the aims were to investigate how patients interpret the questions, which considerations patients had for their specific interpretation of the items, and whether their thoughts and considerations were appropriate for the available response options. In other words, is the questionnaire applicable for use in sub-acute neck pain patients.

**Method**

This qualitative study was conducted within the framework of a randomised clinical trial. Patients were recruited by a general practitioner or a physical therapist, and they were all patients with sub-acute non-specific neck pain, defined as pain in the cervical region existing for at least 4 weeks, but no longer than 12 weeks. The neck pain could radiate to the shoulder region or the upper extremities, or be accompanied by headache, but the main complaint had to concern the neck. Other inclusion criteria were: age between 18 and 70 years, and a new episode of pain (defined as no neck pain in the previous 4 months) and no therapy for neck complaints in the previous 4 months. The exclusion criterion was: specific neck pain, for example due to rheumatoid arthritis, disc herniation, neurological diseases or malignancy.

In this study the Three-Step Test Interview (TSTI) was used. The aim of the TSTI is to identify problems with self-administered questionnaires, and it collects data on how respondents actually complete a questionnaire. The 'think-aloud techniques', that are used, results in a cognitive interview, which consists of the following three steps:

1) Concurrent think-aloud

During this first step the respondent fills in the questionnaire while thinking aloud. Thus, the respondent verbalizes his/her thoughts. This phase is completely observational, which means that no intervention,
questions or comments from the interviewer are allowed, but only encouragement for the patient, if necessary, to think aloud. Thus, the respondent should fill in the questionnaire as if he/she were alone in the room. The interviewer makes notes about the respondent’s behaviour (silences, hesitation, skipping of questions, correction of the response category) and verbalized thoughts. These notes will be used in step 2 and 3 of the TSTI.

2) Retrospective interview

The second step is a more narrative approach, aimed at clarifying the first step (e.g. “I noticed that you hesitated between two response categories, why was that?”). The interviewer only discusses those observed actions or thoughts which he/she feels doubtful. It is important that the respondent reflects on the first step. He or she should not correct an answer in this phase but explain why he/she did or said something in the previous step.

3) Semi-structured interview

The final step is an in-depth interview aimed at eliciting the patient’s considerations and opinions. Additional data is collected, i.e. the patient’s opinion about the questionnaire or the response categories etc, and this information is added to the observational data. It is the only step in the TSTI in which the respondent is allowed, for example, to explain his/her response behaviour, make comments about the items, the questionnaire or give additional information.

During the phase of data collection we used the saturation principle to decide whether additional patients were needed. Patient recruitment stopped if it was expected that no further information should add to the final conclusion. This also meant that in the end of the research the duration of the interviews was diminished, as the in-depth interview focussed on the new information. The research was ended when additional patients did not add further information.

Before the interviews took place the patients were informed about the procedure, and they were asked if they wished to participate. The TSTI, which was held in the patient's home, i.e. close to the natural environment of the patient, took between one and two hours, was audio-taped with a digital voice-recorder, transcribed verbatim and analysed. In the process of analysis, each transcribed interview was divided into three segments, according to the three steps of the TSTI. After fragmentation of the text protocol, the fragments were labelled and categorised, and a table was constructed with the most frequently used labels. The labels were then analysed and described.
Chapter 8

The interviews and analysis were performed by an experienced qualitative researcher (SH).

Results

Process
After 13 patients, 7 women and 6 men, saturation was completed. The average age was 51 years. The think-aloud method was a new experience for all of the patients participating in the research. None of them had ever filled in a questionnaire whilst thinking aloud. For this reason the first phase was difficult for some of the patients. Furthermore, we noticed that some of the patients took their time when filling in the questionnaire, while others answered the questions relatively quickly. The fact that some patients answered the questions quite quickly did not necessarily mean that the questions were more clear to them. The second step of the TSTI provided more insight into their thoughts and considerations. Although the patients did clarify their answers, talking about the items and the response options, often changed their interpretation and this sometimes led to the wish to change their initial response. This could mean that they did not completely understand the items, or misinterpreted the items in the first step. In the third and final step of the TSTI the patients did not hesitate to provide additional information. This step clearly was less problematic than the first two steps. Most remarks concerned the response options and the formulation of a number of questions. One patient even made a comment that from now on he/she will think aloud filling in questionnaires.

Results regarding the TSK
During the interview two main problems were identified. One concerned words within the items, which the patients found difficult to interpret. The two words that most frequently raised problems were “dangerous” and “injury”. The other problem was that implicit assumptions of the questionnaire were interpreted differently by the patients. (See Table 1)

The questionnaire originally was developed in a population of chronic low back pain patients who often consider their complaints as frightening for their activities in daily life. In the current study an important theme was, “are my neck complaints ‘dangerous’ or not”? Most patients disagree with this word and found that the word dangerous was not the appropriate way of describing their health status. In items 3, 8, 11 and 16 the word dangerous or dangerously are used. Patient 3 as an example; "What’s dangerous? You can think of " am I
going to die soon?” or is it an illness that can’t be cured, or something like that. I think it’s rather strongly put.”

The phrase “put my body at risk for the rest of my life” also falls into the same category. Sub-acute neck complaints are not perceived by patients as a risk for severe disability for the rest of their lives. It “hurts but it does no harm” is more likely to be what most patients think. They then find it difficult to choose the appropriate answer, because they feel that “dangerous” is not an issue with regard to their health status.

The second problematic word was “injury”. This word is used in questions 7, 13 and 15. The word injury is ambiguous and the Dutch translation, ‘letsel’, might have different meanings. Some patients interpret this word as “there is something wrong” or “something has been broken”, or “there is a problem” or “there must have been an accident”. Patient 2 said, as an example: “...look, if you think about injury you think perhaps that something is broken or you have been wounded or it’s really bad, but in your neck it is just ‘wear and tear’ and that sort of thing...”.

Another problem is that the implicit assumption made in some items does not apply to the experience or current status of the patients. ‘Physically active’ or ‘exercises’ in the items 12, 13, 14 and 17 were interpreted by patients as meaning that they did not exercise at all or were not physically active. Examples are “I would be better off if I was physically active”. Most patients reacted by saying “I am physically active”. This led to problems in choosing the appropriate response option. In item 10 “unnecessary movements” raised a lot of questions about what exactly unnecessary movements are. Another problem was the word “normal” in item 10, which seemed for some patients with sub-acute neck pain to mean that they were not “normal”. The meanings that patients attribute to the above-mentioned words seem to differ from the meanings attributed by those who developed the questionnaire.

An additional finding in step three was that patients often commented on the lack of a response option “not applicable”.

The applicability of the Tampa scale of Kinesiophobia: a qualitative study
**Table 1:** The Tampa Scale of Kinesiophobia, with identified problems indicated: different words in bold and implicit assumptions in italic, underlined and bold.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I’m afraid that I might injury myself if I exercise</td>
</tr>
<tr>
<td>2.</td>
<td>If I were to try to overcome it, my pain would increase</td>
</tr>
<tr>
<td>3.</td>
<td>My body is telling me I have something <strong>dangerously</strong> wrong</td>
</tr>
<tr>
<td>4.</td>
<td>My pain would probably be relieved if I were to exercise</td>
</tr>
<tr>
<td>5.</td>
<td>People aren’t taking my medical condition seriously enough</td>
</tr>
<tr>
<td>6.</td>
<td>My accident has put my body at risk for the rest of my life</td>
</tr>
<tr>
<td>7.</td>
<td>Pain always means I have <strong>injured</strong> my body</td>
</tr>
<tr>
<td>8.</td>
<td>Just because something aggravates my pain does not mean it is <strong>dangerous</strong></td>
</tr>
<tr>
<td>9.</td>
<td>I am afraid that I might <strong>injure</strong> myself accidentally</td>
</tr>
<tr>
<td>10.</td>
<td>Simply being careful that I do not make any <strong>unnecessary movements</strong> is the safest thing I can do to prevent my pain from worsening</td>
</tr>
<tr>
<td>11.</td>
<td>I wouldn’t have this much pain if there weren’t something potentially <strong>dangerous</strong> going on in my body</td>
</tr>
<tr>
<td>12.</td>
<td>Although my condition is painful, I would be better off if I were <strong>physically active</strong></td>
</tr>
<tr>
<td>13.</td>
<td>Pain lets me know when to stop <strong>exercising</strong> so that I don’t <strong>injure</strong> myself</td>
</tr>
<tr>
<td>14.</td>
<td>It’s really not safe for a person with a condition like mine to be <strong>physically active</strong></td>
</tr>
<tr>
<td>15.</td>
<td>I can’t do all the things normal people do because it’s too easy for me to get <strong>injured</strong></td>
</tr>
<tr>
<td>16.</td>
<td>Even though something is causing me a lot of pain, I don’t think it’s actually <strong>dangerous</strong></td>
</tr>
<tr>
<td>17.</td>
<td>No one should have to <strong>exercise</strong> when he/she is in pain</td>
</tr>
</tbody>
</table>
Discussion

The purpose of this study was to qualitatively evaluate patients understanding and interpretation of the wording used in test items of the TSK. This may be an important missing step in the development of many psychometric questionnaires.

Quantitative research on the Dutch version of the TSK for patients with low back pain showed acceptable validity and reliability. Furthermore, a confirmatory factor analysis suggested a two-factor model, an activity avoidance focus (items 1, 2, 9, 10, 13, 14, 15, and 17) and a pathological somatic focus (items 3, 5, 6, 7, 11). Patients with neck pain appeared to have difficulty with some items when completing the TSK. Qualitative studies of how patients experience and respond to questionnaires can identify these problems, but are still few and far between.

One cause of these problems could be the translation of the questionnaire from English into Dutch. Some correctly translated words can still be interpreted differently. Examples in the TSK are the words “dangerous” and “injury”. In Dutch these might have a different meaning within a certain context. It also is possible that the connotation of the above-mentioned words is different for patients with sub-acute neck pain than for patients with chronic low back pain. The difference between these two groups of patients is the location of the pain and the duration of the pain. In questionnaires it is assumed that the attributions and beliefs of the patients in both pain groups are the same, but it is in fact likely that when completing questionnaires patients with chronic low back pain will have a different subjective feeling about the meaning of words such as “dangerous” or “injury”. Furthermore, the impact of chronic low back pain on activities in daily life is greater than that of sub-acute neck pain. Activities such as sitting, standing and moving around are less affected. However, the extent to which these differences influence the total score and the interpretation of the TSK is not known.

In our study it became clear that some words that are used in the TSK items are troublesome for patients and the question that arises is: should these items be replaced or rephrased? If words are too extreme the patients do not consider these specific words to be appropriate, and find it difficult to choose a suitable answer, should this change the original TSK? Furthermore, fear-avoidance beliefs and pain catastrophizing are already found in a healthy population, and it has been suggested that both factors contribute to the transition from acute to chronic pain. According to Buer and Linton, the localisation of the
pain has limited influence. However, it became clear during our study
that patients suffering from sub-acute neck pain has less pain related
fear than patients suffering from low back pain. It can be hypothesized
that neck pain patients have different beliefs, compared to patients
suffering from chronic low back pain, specially concerning the meaning
of “dangerous” and “injury”.

Another finding concerned criticism about the lack of choice in response
categories, specially with regard to “not applicable”. Because of the
above mentioned difference in attribution some items do not seem to be
appropriate. For example for a patient who is active in sports and
exercise regularly, item 14 is hard to answer with the available response
categories. Adding ‘not applicable’ as an option would make the choice
of categories more suitable.

The advantage of TSTI is its clear structure. It is easy to use, and
although some patients had to be encouraged to think aloud it is
possible to obtain insight into their interpretation of thoughts and
considerations when answering all items of the TSK. In the second step
the interviewer can ask for more information and the patient has an
opportunity to elaborate on the questionnaire and reflect on further
details in the final step. However, this cognitive interview technique
does not include a wider enquiry into people’s lives and problems,
because it specifically aims to adhere closely to the content of the
questionnaire. The only disadvantage of using the TSTI is it’s time
consuming approach. But using the saturation principle it is not
necessary to include a large amount of patients.

It has been suggested that the TSK should be used without the
reversed keys of items 4,8,12 and 16, because the factor analysis
showed a weak association with the total TSK score\footnote{10} and they were not
considered to be appropriate. One could hypothesize that specific words
used in these items were troublesome. In our study the patients had no
specific problems with difficult or ambiguous words or inadequate
assumptions in these reversed key items.

Our study suggests that a ‘patient-centred’ approach such as the TSTI to
improve the formulation of item’s, are an important aspect in the
development and validation of a questionnaire. Furthermore, this study
showed that many of the issues mentioned above can affect the
interpretation of the questionnaire. Patients with sub-acute neck pain
seems to have less pain related fear, possibly due to the duration of the
pain. Neck pain therefore seems to have different psychological consequences. Words such as “dangerous” or “injury” seem to have a different meaning for patients with neck pain than for patients with other chronic pain conditions. To extent to which these problems influence the interpretation of the TSK score is still a subject of debate. Furthermore, the above-mentioned problems have, in our opinion, an effect on the applicability of the TSK, so further research must be carried out to determine whether and, if so how these problems can be addressed. The effect of adaptation of the TSK to other pain conditions also needs to be studied.

In qualitative research the main aim is to collect opinions about for example the interpretation of questionnaires such as the TSK rather than to collect data of a population sample. The principle of saturation is used to ensure that all important meanings and opinions are collected. A representative sample and generalisation to the total population are no issues in qualitative research. When new patients doesn’t add new information the research can end. In the current study after 13 patients the saturation was completed.

**Conclusion**

It is recommended that qualitative research should be carried out to supplement quantitative research in the validation of a psychological questionnaire. It is important that not only all the psychometric properties are assessed but that also interpretations and thoughts about meaning of words and considerations in choosing a response option are important issues in the development and validation of questionnaires.
Reference List


Added value of qualitative studies in the development of Health Related Patient Reported Outcomes

Abstract

Psychometric analyses, such as factor analyses, internal consistency and construct validity, are well known and frequently applied methods in the development of health related patient reported outcomes. These statistical indexes shed hardly any light on how respondents interpret individual items, or on the meaning of their responses. In this study, the Pain Coping and Cognition List (PCCL), a quantitatively validated psychological questionnaire developed for chronic pain, has been subjected to a qualitative research method: the Three Step Test Interview (TSTI), an observational technique that aims to discover problems with self reported questionnaires. It consists of three phases: 1) concurrent think aloud; 2) a retrospective interview; 3) a semi structured interview. Six different types of problems were distinguished: long complicated formulations, composite questions, irrelevant questions, lacking frame of reference, problematic words, and wrongly interpreted questions. This study illustrates that quantitative methods have an added value when developing self reported questionnaires because problems were highlighted that can not be identified using quantitative methods only. Therefore, we recommend that a full qualitative study should be an integral part of the development of questionnaires. The TSTI method is very useful for this purpose.
Introduction

When developing questionnaires it is important that these questionnaires provide valid and reliable results. A prerequisite is that the questionnaires measures what it is supposed to measure. For this, it is important that respondents do understand the items in the questionnaires, and therefore give answers on the individual items as meant by the developers of the questionnaire. In other words, it is easy to fall into the trap of using questionnaires like a form of laboratory equipment and to forget that, like most social research tools, they are open to various interpretations. Psychometric analyses, such as factor analysis, internal consistency and construct validity, are well known and frequently applied methods in the development of health related patient reported outcomes such as questionnaires. These quantitative methods apply statistical techniques to assess the psychometrical characteristics of a questionnaire. In doing so, these techniques will highlight some problems, if present, with the items of a questionnaire, for example in terms of a low item-total correlation, indicating that such an item does not fit into the questionnaire satisfactorily. But these statistical indexes shed hardly any light on how respondents interpret individual items, or on the meaning of their responses. Qualitative research methods are well suited to examine the interpretation of psychological self-administered questionnaires.

In this study, a psychological questionnaire, the Pain Coping and Cognition List (PCCL)\textsuperscript{6}, has been subjected to such a qualitative research method. The PCCL has been developed to measure coping, cognitions and locus of control in patients with chronic pain. It is a self-completion questionnaire and consists of 42 items to be answered on a 6 point Likert scale ranging from ‘I completely disagree’ to ‘I completely agree’. The PCCL is a combination of the Pain Cognition List (developed and validated by Vlaeyen et al\textsuperscript{10}, the Locus of Pain Control Questionnaire (developed by Engstrom\textsuperscript{2} (1983) and translated in Dutch and validated by ter Kuile et al\textsuperscript{7,8} (1993, 1999), and the Coping Strategies Questionnaire (developed by Rosenstiel \& Keefe\textsuperscript{4} and translated in Dutch and validated by Spinhoven et al\textsuperscript{5}). Of these individual questionnaires internal consistency, test-retest reliability and construct validity have been tested, and these psychometric characteristics have shown to be acceptable\textsuperscript{6}. In the pilot phase it was checked whether a small number of patients understood the questions well. Via factor analysis the PCCL was constructed to measure cognitions, coping strategies and locus of control. Although the PCCL and its components has been validated
quantitatively\textsuperscript{5}, little research is done on how patients interpret the various items, and, consequently, how the data that this questionnaire produces should be interpreted.

In an ongoing randomised controlled clinical trial on neck pain\textsuperscript{3}, patients with sub-acute neck pain are asked to complete the PCCL. Although according to the developers, it takes 10-15 minutes to complete, in this study it appeared that most patients need about 15 – 20 minutes to complete the questionnaire, but some patients take an extremely long period thinking. This might be due to the fact that patients have difficulties in understanding the items, and/or that their thoughts and considerations fit poorly into the response categories.

The objectives of this qualitative study were to find out what kind of problems patients with sub-acute neck pain are facing when filling in the PCCL which was originally developed for chronic pain patients, and how patients interpret the questions in the questionnaire. This qualitative study will help to interpret the findings of the PCCL in the quantitative study, but also in previous and future studies. It may also lead to further adaptations of (items in) the questionnaire or more extensive instructions for its use.

**Method**

**Patients**

Patients were recruited by one general practitioner, three physiotherapists, and one manual therapist. They informed the patients about the neck pain research and asked them if they wanted to take part in the Three Step test Interview (TSTI). The patients with non specific neck pain, with a duration of 4 to 12 weeks, were invited to participate in the current study. The TSTI took place at the patient’s home, thus in the natural environment of the patient. This was done for two reasons: firstly, in this way most reliable data are obtained in a qualitative study\textsuperscript{1}, and secondly to collect the qualitative data in the same environment as the randomised controlled clinical trial. The patients were interviewed once-only for about one to two hours. The interviews were recorded with a digital voice recorder and transcribed verbatim.

In a qualitative study, it is more important to register what patients have to say about a certain item than the number of patients stating a
certain opinion. According to the saturation principle, patients were included till no new information (opinions or viewpoints) was brought forward. Saturation occurred after interviewing thirteen patients. Of these 13, the last four patients were asked to complete only those parts of the PCCL talking aloud, that provided new information in the three previous interviews.

The Three Step Test Interview
In the study The Three Step Test Interview (TSTI) \(^9\) was used as research method. The TSTI is an observational technique to discover problems with self-administered questionnaires. It is aimed at collecting data on how respondents actually complete a questionnaire. Because of the use of think aloud techniques, it can be seen as a cognitive interview. The TSTI consists of the following three phases:

1) Concurrent think aloud
During this step the respondent fills in the questionnaire while thinking aloud. Thus, the respondent verbalizes his/her thoughts. This first phase has a completely observational nature, which means that no intervention (questions or comments) by the interviewer is allowed. Thus, the respondent should fill in the questionnaire as if he/she were alone in the room. The interviewer makes notes of the respondent’s behaviour (hesitation, skipping of questions, correction of the response category) and of verbalized thoughts. All verbal and observational information retrieved in this phase are used as the input for step 2 and 3 of the TSTI.

2) Retrospective interview
The second step is aimed at filling in the gaps of the first step and at checking information (e.g. “You hesitated between two response categories, why was that?”). The interviewer considers those observed behaviours, actions or thoughts, which he/she does not feel fully informed about. It is important that the respondent does not report what he/she thinks now, thus in the second step, but tells what his/her thoughts and considerations were in the first step.

3) Semi-structured interview
The final step is an in-depth interview aimed at eliciting the respondent’s considerations and opinions. In this phase additional data is collected which is added to the observational data. It is the only step in the TSTI where the respondent is allowed to explain his/her behaviour, actions or
Chapter 9

thoughts of the previous steps. For example, give comments on the items of the questionnaire or give additional information on the subject.

Two researchers interviewed the participants. One researcher (SH) was the primary interviewer responsible for the analysis and reporting of the qualitative study. A second researcher (JP), an expert in neck pain, was present during all the interviews and sometimes took part in the conversation. This was useful for the data interpretation, which was performed by SH and JP together.

Data synthesis
Before analysing the interviews, Kwalitan, a program of qualitative research, was used to divide each transcribed interview into segments. Each segment contained one of the three phases of the TSTI. Thereafter, the patients’ statements were placed in a table, per item and per phase. Then the type of problems and interpretations were analysed and categorised per item. The problems and interpretations of the items that appeared to be problematic were illustrated with examples of typical citations.

Results

Performance of the Three Step Test Interview
The think aloud method was a new experience to all of the participants in the current study as none of them had ever filled in a questionnaire ‘thinking aloud’. For this reason the first phase appeared to be difficult for some patients. A number of older patients clearly faced more difficulty than some of the younger patients. Moreover, patients with a higher education appeared to be more critical with regards to the formulation of the questions. Moreover, we noticed that some patients took their time when filling in the questionnaire, while others answered the questions relatively quickly. The fact that these patients answered the questions rather quickly did not necessarily mean that the questions were more clear to them. In comparison to step one, the second step gave more insight into the thoughts and considerations of the patients. Although patients did clarify their answer to the questions and gaps were filled in, it appeared that some patients would have answered some of the questions differently in the second phase. This could mean that they did not completely understand or misinterpreted the questions in the first phase. In the third and final phase of the TSTI patients freely gave additional information. This phase clearly was much less problematic than phase one and two. Most opinions concerned
comments on the response category and the formulation of a number of questions.

The questions
In the analysis it appeared that 18 of the 42 items of the PCCL were more or less problematic for the patients. These items were unclear, difficult to interpret or wrongly interpreted by the patients. Six categories of types of problems were identified: complicated formulations, composite questions, irrelevant questions; questions lacking a clear frame of reference; questions containing problematic words; wrongly interpreted questions. For some of these categories a typical example is presented.

1) long complicated questions (item 16, 31):
Needles to say: questions that are (too) long or complicated may be problematic to answer for respondents. This is illustrated in the following example:
Item 16: “I find that the pain bothers me less if I do all the things that I usually do to cope with my pain.”
Patient 2: “Well, you shouldn’t spread this sentence out too much, because it’s too complicated and too long.”
It appeared to be one of the most problematic items of the PCCL. Patients had to read the question several times in order to understand the intention of the question, mainly due to the length and complexity of the item. But even then it was still uncertain whether the item was completely understood.

2) composite questions (item 2, 24, 27, 35):
If a questions contains actually two or more questions it is clear that to answer such a question is problematic. Item 24 is an example of a composite question.
Item 24: “When I have pain I feel terrible and I feel that it’s all too much for me.”
Patient 10: Then I feel terrible, and I have the feeling that it’s too much for me. Well, wait a minute, that last bit, you see, it’s really two separate parts. Which question do I have to answer here? Do I feel terrible? Or do I feel that it’s too much for me? ... You can’t really answer all that in one question, or can you? ... also, I don’t really see the connection...you can... you can feel terrible, but that doesn’t mean that you feel it’s all too much.”
In addition to item 24, also item 2, 27 and 35 of the PCCL consists of composite questions. As illustrated in the quote above, these type of
questions run the risk of being answered partly by respondents. Moreover, these questions contain sub questions with often different intentions. This patient clearly did not see the connection between the two sub questions.

3) Irrelevant questions (item 5, 25):
The third problem concerns irrelevant questions, meaning that they are not applicable to part of the patients.
Item 25: “Relaxing exercises reduce the pain.”
Patient 7: “I haven’t .... I haven't experienced that yet. I haven't done any relaxing exercises. So I'm sorry, I can't answer that.”
It is clear that this item only appeared to be relevant to patients who actually do, or have the experience of, relaxing exercises. This patient had never done these kind of exercises. Therefore it was difficult for him/her to answer the question.

4) Frame of reference is lacking (item 4, 20):
Some questions lack a clear frame of reference. The patients need a frame of reference with regard to time period, pain, and activities. For example, some of the patients did not know whether to base their answer on the present situation or on the beginning of the episode of pain, when the pain was worse.

5) Problematic words (item 1, 13, 15, 19, and 38):
Some questions contain problematic words. The following words appeared to be problematic: “positive” (item 1), “daily life” (item 13), “doctors” (item 15), “complete human being” (item 19), “relax physically” (item 38).
Item 38: “The pain is less if I relax my body.”
Patient 2: (...) Yes but then you have to ....um....relax your body – does that mean sitting still, or is that...um ....walking or cycling...?”
According to this patient “relax physically” is multi interpretable. It is unclear to him/her what is exactly meant by these words.

6) wrongly interpreted questions (3, 21, 36):
Three of the items in the PCCL were interpreted in another way than the developers intended.
Item 3: “When I have pain I pray that the pain will stop. No, I’m not religious and I don’t pray. Yes, you feel like doing that, but... I think that’s so unimaginable, so I don’t agree. Totally disagree.”
Nearly all patients based their answer on whether they are religious or not. However, this is not the intention of the item. The meaning of the
Added value of qualitative studies in the development of outcomes

Item is to see how patients deal with their pain externally. Only one patient interpreted the item in the way it is intended:

Item 3: “When I have pain I pray that the pain will stop.”

Patient 2: “Well, you know, praying is like begging, or like I hope that the pain will stop. That seems to me to be a normal attitude, if you hope that the pain will pass, so I agree with that.”

Patient 2 interprets “praying” as “begging” or as “hoping that the pain stops”. The patient does not approach the item on a religious level. As far as the other patients are concerned, there clearly seems to be an interpretation problem. Item 3 has an interpretation problem. Nearly all the patients base their answer on the fact that they are not religious, but the intention of the item is not to find out whether the patient is religious or not, but to see whether a patient seeks help from others to reduce their pain, which is a typical example of passive coping. Table 1 gives an overview of the types of problems identified.

Table 1: Overview of the type of problems identified in the PCCL items

<table>
<thead>
<tr>
<th>Type of problem</th>
<th>Item number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long and complicated items</td>
<td>16, 31</td>
</tr>
<tr>
<td>Composite items</td>
<td>2, 24, 27, 35</td>
</tr>
<tr>
<td>Irrelevant items</td>
<td>5, 25</td>
</tr>
<tr>
<td>Lacking frame of reference</td>
<td>4, 20</td>
</tr>
<tr>
<td>Containing problematic words</td>
<td>1, 13, 15, 19, 38</td>
</tr>
<tr>
<td>Wrongly interpreted items</td>
<td>3, 21, 36</td>
</tr>
</tbody>
</table>

Use of answering categories

In addition to the five types of problems with respect to the questions, some problems concerned the somewhat contradictory way of patients to come to a certain response category.

Item 4: “To avoid more pain I have to make regular visits to a doctor or some other care-provider.”

Patient 1: (…)Yes, then I want to hear whether I only have to think about this particular bout of neck pain now…. I don’t usually make regular visits to a doctor or a care-provider, but now I’ll have to, because it won’t just disappear. So, I don’t totally agree with that.”

This patient normally does not go to a doctor or another person for treatment on a regular basis. However, his/her pain does not go away, and therefore he/she does need to see a doctor. Given his/her thoughts,
it is remarkable that the patient rather disagrees with the item. Another example which illustrates this case is:
Item 27: "When I have pain I tell myself that I must not let myself be hindered by the pain and do what I want to do."
Patient 13: “Actually, I don’t so much say it to myself, it’s more like an attitude that I always have....um.....I don’t go and lie down in my nest like a little bird, so I somewhat agree.”
This patient also fills in a complete opposite response category than expected.
There were also patients who just filled in an answer, even though they did not understand the item at all.
Item 16: “I find that the pain bothers me less if I do all the things that I usually do to cope with my pain.”
Patient 13: “Wait a minute, I’m going to read that question again.... I don’t understand the question. I find it ..... it’s difficult to understand....um....Yeah, you have to guess at an answer. I don’t really understand what....what the meaning of the question is. So, to be safe I would say that I slightly agree.”
Obviously, patient 13 did not understand the intention of the item. After reading the item a second time he/she still did not understand the question. Finally, the patient just guessed an answer.
Not surprisingly, patients who guessed an answer all filled in a response category that is somewhat in the middle of the scale.

Discussion

Although the PCCL and its components has been validated quantitatively\(^2\), this study illustrates that qualitative methods identify problems that can not be retrieved by quantitative methods, such as factor analysis or internal consistency. Based on these statistical techniques these items were not excluded from the questionnaire. But still 43% of the items (18 out of the 42!) raised problems to a greater or a lesser extent. The TSTI proved to be a useful method in finding out what problems patients with sub-acute neck pain are facing when filling in a psychological questionnaire and how they interpret the items.

This study yields rather convincing evidence that the length and complexity of items 16 and 31 raise many problems. For some patients the items even remain unclear after reading them several times. Others just filled in an answer, even though they did not understand the item(s) at all. This of course hampers the validity of these items and leads to unreliable results, which should be prevented.
There are several items (2, 24, 27, and 35) that contain composite questions. As can be expected, these items, despite the fact that they were included after quantitative methods, raise problems. The most prominent problem of these items is that they are at risk of being partly answered. Needless to say, one does not know to which part of the question (or sub question) the answer applies.

An important question is whether we should change these problematic items. We know that it is a challenge to change the items, without modifying the meaning of the question. If the questionnaire is used frequently and by various researchers and among different populations, this is not the most obvious solution as different versions of a questionnaire could make it difficult to compare research results. On the other hand, if a questionnaire is not reliable nor valid the questionnaire should not be used at all. The questionnaire we used as an illustration in this study, the PCCL, is not used often yet, so changing formulations is defensible in this case. This discussion of changing specific items, or not, in existing questionnaires emphasize the importance of eliciting these kind of problems beforehand, i.e. during the development of the questionnaire. And although this is certainly not a landslide breakthrough, this study illustrates that it is not uncommon to develop a questionnaire mainly (or solely) based on quantitative techniques, thereby missing the opportunity to evaluate the meaning and the interpretation of a questionnaire.

This does not only hold for developing a new questionnaire. Often questionnaires developed (and quantitatively tested) in a given population are applied in other (seemingly resemble) the original population. Our example, the PCCL, has been developed for patients with chronic pain. In the current study we applied the PCCL to a population of patients with sub-acute pain, whose pain is not present that long and in many cases the intensity of the pain is less severe. Some of the problems encountered, especially the items which were considered irrelevant or lacked a frame of reference, were attributable to the application in a new target population. The pain in these patients is not that severe or of sufficient long duration to answer the items correctly. An answering category ‘not applicable’ might solve this problem. Some questions were formulated far too strong according to sub-acute pain patients. But actually they were intended that way by the developers, focussing on chronic pain. Therefore it can be questioned
whether these questions should be adapted, or that this questionnaire should not be used, as it is, in this population. In other words, this illustrates that also when a questionnaire is developed in ‘adjacent’ populations qualitative methods are an important step not to be omitted.

Conclusion

Quantitative methods, such as the TSTI have an important added value, in addition to quantitative methods in the development of questionnaires. The current study illustrates that in a questionnaire that has been validated quantitatively, by factor analysis, internal consistency and construct validity and even after a small number of patients have checked the questionnaire on clarity, still a considerable number of items of the questionnaire were problematic for the patients. Therefore we recommend that quantitative methods should be an integral part in the development of new questionnaires. The TSTI method is very useful for this purpose.
Reference List

Chapter 9

Appendix

English version of the Pain Coping and Cognition List (PCCL)

1. I think I could have a positive influence on my pain.
2. When I have pain I see it as a challenge, and I don’t let it get me down.
3. When I have pain I pray that the pain will stop.
4. I get less pain if I think about nice things.
5. To avoid having more pain I have to make regular visits to a doctor or some other care provider.
6. When I have pain I go to see other people.
7. The pain gets less when I think about things that are happening all around me.
8. Whatever I do, I can’t change my pain in any way.
9. When I have pain I act as if the pain is not part of me.
10. I think that I am active and busy.
11. When I have pain I act as if the pain is not there.
12. As far as my pain is concerned, I can only do what the doctor or some other care-provider says.
13. I think that I can look after myself very well in daily life.
14. I know a way in which I can reduce my pain to a certain extent.
15. Only the doctors can help me with my pain.
16. I find that the pain bothers me less if I do all the things that I usually do to cope with my pain.
17. When I have pain I tell myself that I can conquer the pain.
18. I think I’m a hopeless case.
19. Because of my pain I no longer feel like a complete person.
20. I’ve become physically much weaker.
21. When I have pain I lose my faith in God.
22. It seems as if my pain is becoming more important.
23. When I have pain I do everything I can to avoid thinking about the pain.
24. When I have pain I feel terrible, and I feel that it’s all too much for me.
25. Relaxing exercises reduce the pain.
26. If I get more pain I have to consult my doctor or another care-provider.
27. When I have pain I tell myself that I must not let myself be hindered by the pain and that I must do what I want to do.
28. I feel that I’m able to do less all the time.
29. Because of the pain I no longer get around to doing lots of things.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>When I have pain I think about people I like to do things with.</td>
</tr>
<tr>
<td>31</td>
<td>I can reduce the severity of the pain if I go and do all the things that I usually do to cope with my pain.</td>
</tr>
<tr>
<td>32</td>
<td>The word pain makes me afraid.</td>
</tr>
<tr>
<td>33</td>
<td>It depends on me myself, how much influence the pain has on me.</td>
</tr>
<tr>
<td>34</td>
<td>I can reduce the pain by not thinking about it.</td>
</tr>
<tr>
<td>35</td>
<td>When I have pain I know that there will be somebody to help me and that the pain will go away for a little while.</td>
</tr>
<tr>
<td>36</td>
<td>When I have pain I pray that it won’t last much longer.</td>
</tr>
<tr>
<td>37</td>
<td>When I have pain I ignore the pain.</td>
</tr>
<tr>
<td>38</td>
<td>The pain is less if I relax my body.</td>
</tr>
<tr>
<td>39</td>
<td>Because of my attitude I feel I can cope with my pain.</td>
</tr>
<tr>
<td>40</td>
<td>I’m gradually feeling mentally weaker.</td>
</tr>
<tr>
<td>41</td>
<td>When I have pain I think about things that I like to do.</td>
</tr>
<tr>
<td>42</td>
<td>When I have pain I try to think about something nice.</td>
</tr>
</tbody>
</table>
Minimal Clinically Important Change of the Neck Disability Index and the Numerical Rating Scale for Patients with Neck Pain

Submitted as;
Jan J.M. Pool, Raymond W.J.G. Ostelo, Jan L. Hoving, Lex M. Bouter, Henrica C.W. de Vet, Minimal Clinically Important Change of the Neck Disability Index and the Numerical Rating Scale for Patients with Neck Pain
Abstract

Study design. Prospective, single-cohort study.
Objective. To assess the minimal clinically important change (MCIC) on the Neck Disability Index (NDI) and the Numeric Rating Scale (NRS) for pain in patients with neck pain.
Summary of Background Data. Both measurement instruments are frequently used in research and clinical practice, but which changes are clinically relevant is still unknown.
Methods. The MCIC was estimated with two different methods, both integrating an anchor-based and distribution-based approach: the minimal detectable change (MDC) and the optimal cut-off point of the ROC curve. The study population consisted of 183 patients with non-specific neck pain.
Results. The results show an MDC of 10.5 points for the NDI (scale range 0-50) and 4.3 points for the NRS (scale range 0-10), and optimal cut-off points of the ROC curve of 3.5 for the NDI and 2.5 for the NRS.
Conclusions. The estimated MCIC should be used as an indication for relevant changes in clinical practice. Using the optimal cut-off point of the ROC curve, false positives and false negatives are equally weighted and if there are no objections doing so, the optimal cut-off point of the ROC curve may be a good choice.
**Introduction**

Neck pain is a common musculoskeletal disorder, and its point prevalence in the general population of the Netherlands varies between 9% and 22%. Approximately one-third of all adults will experience neck pain during the course of 1 year. Although neck pain is often self-limiting within a few weeks, 40% of the patients contact their general practitioner (GP). Of these, 30% are referred for further diagnosis to a medical specialist and 32% to physiotherapy, manual therapy or some other type of conservative therapy. To evaluate the effect of treatment for neck disorders it is necessary to assess relevant outcome measures, such as pain and functional disability.

The Neck Pain Disability Index (NDI) is a questionnaire that is commonly used in clinical trials to measure the functional status of patients with neck pain. The NDI was originally developed for assessing the functional status of patients with disabling neck pain, particularly whiplash associated disorders. The psychometric properties of the NDI, in terms of validity and reproducibility, is still a topic of research, which also counts for how to interpret change scores.

Vernon et al. assessed face validity through peer-review and patient feedback sessions and concurrent validity of the NDI on the Visual Analogue Scale (n=10 and a correlation of 0.60) and the McGill Pain Questionnaire (n=30 and a correlation of 0.69). Furthermore a test-retest reliability was calculated and found a correlation of 0.89. Hoving et al. assessed the construct and content validity of the NDI using 71 patients with whiplash associated disorders, comparing the NDI with a patient preference questionnaire (PET), the correlation was 0.57 with the remark that the PET identified more disabilities.

In the review of Pietrobon et al., the NDI was found to be one-dimensional, the validity was established by concurrent criterion validity and showed a correlation coefficient of 0.6 with the VAS and 0.7 with the McGill Pain Questionnaire and was reported to be the scale which was most widely validated among different patient populations, the responsiveness was not reported. The Numeric Rating Scale (NRS) is frequently used to measure pain intensity. Patients are asked to rate their pain on a 0 to 10 point rating scale. Bolton et al. compared the responsiveness of three pain scales, Visual Analogue Scale, the Verbal Rating Scale and the Numeric Rating Scale on patients, n=79, and using effects sizes. The NRS showed to be the most responsive (effect size 0.86).

For the interpretation of treatment effects it is not only important to know whether results are statistically significant, but also whether they...
are relevant for patients or clinicians. Consequently, insight into the clinically important difference or change is needed. A well accepted definition of Minimal Clinically Important Difference (MCID) has been proposed by Jaeschke et al\textsuperscript{13} as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in patient’s management”. We prefer to use the term Minimal Clinically Important Change (MCIC) for the change in health status \textit{within} patients and the term MCID to indicate differences \textit{between} patients. The aim of this study is to assess the Minimal Clinically Important Change of both the NDI and the NRS for pain in patients with neck pain. A number of different methods have been proposed to determine the MCIC\textsuperscript{18}. Crosby et al, distinguish an anchor-based and a distribution based method. Anchor-based approaches use an external criterion to operationalise clinically important change, and distribution-based approaches are based on statistical characteristics of the sample, for example effect sizes, relating observed change to the sample variation\textsuperscript{6}. For neck pain effect sizes and standard response mean (SRM) have been used\textsuperscript{3;23}. The estimate of the effect size and of the SRM are parameters without any dimension which makes it difficult to interpret them for clinicians. Therefore we used two methods, both integrating an anchor-based and a distribution-based approach: the Minimal Detectable Change (MDC) and the optimal cut-off point of the Receiver Operator Characteristic Curve (ROC)\textsuperscript{19}, because these are expressed in scale points which improves the interpretability of change scores. For both methods the global perceived effect (GPE) is used as an external criterion for change\textsuperscript{11}.

\textbf{Methods}

\textbf{Study population}

The study population consisted of participants included in a randomised controlled trial to evaluate the effectiveness of three conservative treatment options for neck pain\textsuperscript{11}. General practitioners (n = 42) referred patients with neck pain to one of the four research centers for study selection. The eligibility criteria were: age between 18 and 70 years, pain and/or stiffness in the neck for at least 2 weeks, neck symptoms reproducible during physical examination, willingness to adhere to treatment and measurement regimens, and no physical therapy or manual therapy for neck pain during the previous 6 months. The participants were randomly allocated to either physiotherapy, manual therapy or continued care provided by a general practitioner.
Data were collected at the research centers at baseline, after 7 weeks and after 52 weeks of follow-up, for the present analysis only measurements at baseline and after 7 weeks of follow-up were used. Approval was obtained from the medical ethics committee of the VU University medical center, Amsterdam.

Measurement instruments
The NDI consists of ten items addressing functional activities, such as personal care, lifting, reading, work, driving or cycling, sleeping and recreational activities, and a number of symptoms such as pain intensity, concentration and headache. For each item answering options range from 0 = no disability to 5 = total disability, resulting in a total range of scores from 0 – 50 points.

The NRS is an 11-point rating scale for pain in which 0 = no pain and 10 = worst pain imaginable. Patients were asked to rate their average pain in the previous week.

To assess the global perceived effect (GPE) the patients rated this on a 6-point Likert scale ranging from 1 = completely recovered, to 6 = much worse. We trichotomized this scale: patients who indicated that they were ‘much worse’ were labeled as ‘importantly deteriorated’; patients who indicated that they were ‘slightly improved’, ‘no change’, or ‘slightly worse’, were labeled as ‘not importantly changed’; and were consequently considered not to have experienced an important or clinically relevant improvement or deterioration; patients who indicated that they were ‘completely recovered’ or ‘much improved’ were labeled as ‘importantly improved’. The distributions of these subgroups (labels) were used to estimate the MCIC thereby integrating anchor based and distribution based methods.

Data-analysis
We defined the Minimal Detectable Change (MDC) as the smallest difference in a score that can be detected, considering the variation in changes on the NDI and the NRS observed in persons who were not importantly changed on the external criterion. To determine the MDC, first the standard error of measurement (SEM) was assessed. The SEM indicates the precision of outcome measure and was estimated by taking the square root of the within-subject variance of patients categorized as ‘not importantly changed’ on the GPE. To be 95% confident that observed change is real change and not caused by measurements error, the MDC was calculated as $1.96\times\sqrt{2}\times$ SEM. Observed change is a result of two measurements, baseline and follow-up and therefore occur twice, hence $\sqrt{2}$. Changes greater than the
MDC are consequently considered to indicate real change\(^1\):\(^{15}\):\(^{19}\), because only ‘not importantly improved’ patients were assessed.

The optimal cut-off point of the ROC curve considers the NDI and the NRS as a diagnostic test for discriminating between ‘importantly improved’ and ‘not importantly improved’ patients. The external anchor (GPE) functions as the gold standard, and distinguishes those patients who showed a clinically important change from those who did not. The diagnostic accuracy of a measurement instrument can thus be expressed in terms of sensitivity and specificity for clinically important change, and can be depicted in a ROC curve. The ROC is a graph of the percentage of true positive values (sensitivity) versus the percentage of false positive values (1-specificity) for each possible cut-off change score of the NDI and the NRS. The optimal cut-off point was chosen in such a way that the overall misclassification, i.e., the sum of the percentages of false positive and false negative outcomes, was minimized. False positive outcomes are persons who are ‘not importantly changed’ according to the GPE, but show a change that is greater than the cut-off value on the measurement instrument. False negatives are persons who are ‘importantly improved’ on the GPE, but show less change than the cut-off value on the measurement instrument. For all statistics, SPSS 12 for Windows was used.

Results
Patient characteristics
During a period of 22 months a total of 183 patients with non specific neck pain were included, of who completed the 7-week follow-up. The mean age of the patients was 45.8 years and 60.8% was female. Table 1 shows the characteristics of the participants at baseline.
Table 1: Characteristics of patients (N=183)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years, sd)</td>
<td>45.8 (11.6)</td>
</tr>
<tr>
<td>Female (%)</td>
<td>60.8</td>
</tr>
<tr>
<td>Previous episodes of neck complaints (%)</td>
<td>64.8</td>
</tr>
<tr>
<td>Duration of current episode (%):</td>
<td></td>
</tr>
<tr>
<td>- 2-6 weeks</td>
<td>48.0</td>
</tr>
<tr>
<td>- 7-12 weeks</td>
<td>26.1</td>
</tr>
<tr>
<td>- ≥ 13 weeks</td>
<td>25.9</td>
</tr>
<tr>
<td>Mean Pain Score (sd) †</td>
<td>6.0 (1.9)</td>
</tr>
<tr>
<td>Mean NDI score (sd) ‡</td>
<td>14.5 (7.0)</td>
</tr>
<tr>
<td>Work status employed (%)</td>
<td>73.8</td>
</tr>
</tbody>
</table>

† Numeric Rating Scale ranging from 0 (no pain) to 10 (worst pain)
‡ Neck Disability Index, 10 items ranging from 0-5 points; maximal disability 50 points

Table 2 shows the mean scores and standard deviations at baseline and after 7 week of follow-up for subjects in each of the six categories of the GPE, and for the combined categories, as used in the analysis. After 7 weeks, 94 patients were labeled as ‘importantly improved’, i.e. completely recovered or were much improved. Only 2 patients were deteriorated, so due to the small numbers, they were excluded from the analysis. 87 Patients were labeled as ‘not importantly changed’, i.e. unchanged, slightly improved, or slightly worse.
Table 2: Mean scores and standard deviations of the Neck Disability Index (NDI) and the Numeric Rating Scale for pain (NRS) at baseline (T0) and after 7 weeks (T7) for categories of global perceived effect (GPE).

<table>
<thead>
<tr>
<th>Categories of GPE</th>
<th>NDI n=183</th>
<th>NRS n=182</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T0</td>
<td>T7</td>
</tr>
<tr>
<td>Completely recovered n=19</td>
<td>10.74(5.32)</td>
<td>0.68(1.64)</td>
</tr>
<tr>
<td>Much improved n=75</td>
<td>14.67(7.57)</td>
<td>6.16(4.69)</td>
</tr>
<tr>
<td>Slightly improved (NDI n=48 NRS n=47)</td>
<td>15.71(6.96)</td>
<td>12.00(7.20)</td>
</tr>
<tr>
<td>No change n=29</td>
<td>14.52(6.60)</td>
<td>14.24(6.65)</td>
</tr>
<tr>
<td>Slightly worse n=10</td>
<td>13.50(5.34)</td>
<td>14.10(7.98)</td>
</tr>
<tr>
<td>Much worse n=2</td>
<td>19.00(5.57)</td>
<td>26.50(13.44)</td>
</tr>
<tr>
<td>Total</td>
<td>14.49(7.00)</td>
<td>9.06(7.46)</td>
</tr>
<tr>
<td>Importantly changed * n=94</td>
<td>13.87(7.32)</td>
<td>5.05(4.78)</td>
</tr>
<tr>
<td>Not importantly changed ‡ (NDI n=87 NRS n=86)</td>
<td>15.06(6.65)</td>
<td>12.99(7.11)</td>
</tr>
</tbody>
</table>

*categories “completely recovered” and “much improved” were considered to indicate importantly changed
‡ categories “slightly improved”, “unchanged” and “slightly worse” were considered to indicate “not importantly changed”
Table 3: Minimal detectable change and several possible cut off scores of the ROC curve for the NDI and the NRS

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Range</th>
<th>MDC ‡</th>
<th>ROC cut-off*</th>
<th>Se˚</th>
<th>Spŧ</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDI</td>
<td>0-50</td>
<td>10,5</td>
<td>10.5</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>3.5</strong></td>
<td><strong>0.9</strong></td>
<td><strong>0.7</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>NRS</td>
<td>0-10</td>
<td>4,3</td>
<td>4.5</td>
<td>0.4</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>2.5</strong></td>
<td><strong>0.8</strong></td>
<td><strong>0.8</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
<td>0.9</td>
<td>0.6</td>
</tr>
</tbody>
</table>

‡ MDC, the minimal detectable change
* optimal cut-off point of the ROC curve for clinically important change with a sensitivity ˚ and a specificity ť of the Neck Disability Index and Numeric Rating Scale for pain.

Table 3 presents the MDC and the optimal cut-off point for the stable subjects. The MDC for the NDI is 10.5, and the optimal cut-off point of the ROC curve is 3.5. The MDC for the NRS is 4.3, and the optimal cut-off point of the ROC curve is 2.5. The optimal cut-off point of the ROC curve for the NDI corresponds to a sensitivity of 0.9 and a specificity of 0.7. For the NRS for pain the sensitivity and specificity both were 0.8.
Discussion

In research and also in clinical practice the NDI and the NRS are often used as questionnaires to evaluate the effects of interventions on functional status and pain perception in patients with neck pain. Hence it is important to know what the smallest change in score is on both questionnaires which patients and clinicians label as clinically important. This study demonstrates quite a difference between the two methods used to estimate the MCIC for the NDI as well as the NRS. Using the optimal cut-off point of the ROC curve as a method, both improved and unchanged patients are included. The optimal cut-off point of the ROC curve is chosen in such a way that the percentages of false positive and false negative outcomes are minimized. So, if one wants to weight false positive and false negative misclassifications equally, the optimal cut-off point of the ROC curve is preferred. If one hesitates to classify patients as ‘improved’, of whom the change scores fall within the measurement error of the unchanged patients, one may prefer the more conservative MDC method. The choice between the two methods may depend on the type of intervention or the clinical consequences of being ‘false positive’ or being ‘false negative’.

The MDC for the NDI is 10.5 points on a scale of 50 points. This means for example, that each item of the questionnaire should improve one step on the 6-point Likert scale. This MDC can be considered as quite large, since this magnitude is greater than the change score of patients who consider themselves as ‘much improved’ (mean NDI score = 8.82). So in other words the MDC considers nearly all patients as being within the measurement error of the questionnaire. As a consequence if applying this MDC as a cut off point in clinical practice and the change score is more than the MDC, one knows almost for sure (with a uncertainty of 5%), that a patient really is changed. A possible explanation for this large MDC for the NDI can be that ‘slightly improved’ patients are included in the “unchanged” group and not in the “improved” group. However the inclusion of this group into the ‘unchanged’ group is sensible and has been demonstrated before. Another remark has to be made. Based on these analyses it is unclear whether for deterioration a similar value applies. Despite the small number (n=2) we could not make a estimation of the ‘minimal important deterioration’. However, based on Farrar’s study there is some evidence that patients interpret deterioration quite differently to improvement.
However, the disadvantage of the MDC method is that the false negative rate is not taken into account, in other words if there is no reason for weighting false negative different from false positive we recommend the use of the optimal cut-off point of the ROC curve. When we used the method of the optimal cut-off point of the ROC curve the MCIC of the NDI appeared to be smaller: 3.5 points. This score makes more clinical sense because this change is quit similar to the change in score for those patients who consider themselves as ‘slightly improved’. Since both “improved” and “unchanged” patients are included in the analysis of the cut-off point of the ROC curve, this method anticipates to a more clinical perspective on the change of the questionnaire. However, their still is a possibility of false positive outcomes.

The MDC for the NRS for pain is also quite large: 4.3 points. Again this magnitude equals the mean change score of the patients who consider themselves as much improved. The optimal cut off point of the ROC curve is 2.5 points. This MCIC for pain is in line with the findings of other studies \(^9;19\) in which the ROC curve was used to define clinically important change and in which an average reduction of two points was found to be clinically important.

In summary, we consider the optimal cut-off point of the ROC curve as the most optimal method, since false positive and false negative can be weighted equally.

Although the GPE is often applied as an external anchor, the use of this scale has been criticized by Norman et al\(^14\). They question the validity of a single-item design compared to a multi-item scale. Another disadvantage of the GPE is that it may be difficult for patients to recall their initial health status and to compare it with their current status in order to assess any changes, and this may introduce bias. Fritz and Irrgang found that a global rating of change could be used to differentiate unchanged patients from improved patients in the dimension of physical impairment\(^10\). In line with previous studies\(^2;9;15;19\) we used a GPE scale to cover the whole range from severely deteriorated, slightly deteriorated, no change, slightly improved, much improved, and completely recovered. To calculate of the MCIC, the cut-off point for clinically important change was set at ‘much improved’. So therefore the category ‘slightly improved’ was labelled as ‘not importantly changed’. We had several reasons for this. Firstly, in our opinion this more accurately reflects the concept of clinically important change. Setting the cut-off point for improvement at slightly improved may reflect more accurately the smallest detectable change, and not the

183
minimal clinically important change. Secondly, we think that patients are likely to give ‘slightly improved’ as a socially desirable answer, even if they did not perceive a relevant improvement. Finally, in previous studies\textsuperscript{9,15,19} it was found that the difference between the categories ‘no change’ and ‘slightly improved’ was small whereas the difference between ‘slightly improved’ and ‘much improved’ was greater. These results were confirmed in the present study.

**Conclusion**

The estimated MCIC should be used as an indication for relevant changes in clinical practice. Using the optimal cut-off point of the ROC curve, false positives and false negatives are equally weighted and if there are no objections doing so, the optimal cut-off point of the ROC curve may be a good choice. However, if there are objections against classifying as improved those patients whose results fall within the measurement error of the ‘unchanged’ patients, the more conservative MDC method would be more appropriate.
Reference List


Introduction

The central focus of this thesis is neck pain and disability, a complaint which is frequently encountered in general practice. In most cases the patient will recover within one to two weeks, but sometimes the complaints last longer and eventually become a chronic pain disorder. If the non-specific neck pain continues, the general practitioner (GP) has to decide whether or not the patient should be referred to a physical therapist or a manual therapist. However, this decision is still based on limited scientific evidence.

The first aim of this thesis was to compare the effectiveness of two interventions for patients with non-specific sub-acute neck pain: a behavioural graded activity (BGA) programme and manual therapy (MT). The second aim of the thesis was to assess which factors influence the short and long-term course of the neck pain. As a third aim, we assessed the reliability, validity and interpretation of scores on frequently used clinical tests and questionnaires, focusing on neck pain, neck function and related disabilities.

In this chapter we will discuss the results of the randomised clinical trial (RCT) and the prognostic study, and the outcome of the clinimetric studies. Furthermore, after drawing final conclusions, recommendations will be made for daily clinical practice and future research.

Randomised clinical trial

Background

We compared the effectiveness of a BGA programme and MT in patients with sub-acute non-specific neck pain. The decision to compare these two interventions was based on the fact that: a) a BGA programme was reported to be potentially effective for patients with pain\textsuperscript{25,26} and b) MT was found to have an effect on pain, perceived recovery and disability in patients with neck pain. A recently conducted RCT, performed by Hoving et al. demonstrated these effects in favour of MT compared to physical therapy and usual care provided by the GP\textsuperscript{15,18,23}. MT in their trial consisted mainly of specific mobilisation techniques, sometimes with additional exercises. Furthermore, the cost-effectiveness evaluation in their study demonstrated a difference in costs in favour of MT. In line with this outcome, several systematic reviews\textsuperscript{1,11,17} showed an effect in favour of a combination of exercise and manipulations and/or mobilisation techniques. There was no strong evidence for the
effectiveness of any other specific therapy in the treatment of non-
specific neck pain\textsuperscript{19,12,24}. As in the Hoving trial, the content of the MT in
the present trial consisted of a combination of specific mobilisation
techniques and exercises.

In the past decade, in the field of physical therapy in the Netherlands
increasing attention has been paid to behavioural aspects in the
treatment of patients with pain. These aspects can also influence the
transition from acute to chronic pain\textsuperscript{10,25}, a transition that takes place
after approximately 8 weeks from onset. A BGA programme is designed
to focus specifically on these behavioural aspects\textsuperscript{21}. The effectiveness of
a BGA programme has been studied before in patients with chronic low
back pain\textsuperscript{29,36}, osteoarthritis of the knee or hip\textsuperscript{39}, and peripartum pelvic
pain\textsuperscript{5}. We hypothesised that these behavioural aspects would also
influence the course of pain in patients with neck pain, and that a BGA
programme might therefore be effective for these patients.

We decided to conduct an RCT which compared the effects of a BGA
programme to the effects of MT. Taking into account the fact that in
the Hoving trial 48\% of the patients had sub-acute neck pain, and that
psychological factors have been reported to effect the transition from
sub-acute to chronic pain\textsuperscript{25,26}, we focussed on this specific sub-group.
The outcomes of our study we will be discussed in a critical review of the
RCT, focussing on the interventions and the role of the GPs, the care-
providers and the patients.

\textbf{Main findings}

- In treatment of patients with sub-acute neck pain, we found a
  marginally significant difference in favour of the BGA programme
  on the outcome disability in both the short term and the long
term (Chapter 3).
Chapter 11

Discussion

Behavioural graded activity programme

In the international literature on physiotherapy in a primary care setting there is no detailed description of BGA programmes. The focus of a behavioural intervention in primary care is on increasing a patient's level of activity- and/or participation. Advice that is often given to increase the patient's level of activity is: 'stay active', 'pain hurts but does no harm' or 'keep moving'. Such advice is based on the biopsychosocial model which infers that not only physical factors, but also psychosocial factors cause the pain. To perform an adequate BGA programme the attitude and beliefs of the therapist must be based on this biopsychosocial model instead of a biomedical model. In the latter model an increase in the patient's pain indicates more damage, and might lead to advices such as `take more rest when you feel the pain`.

The question that must be addressed is whether physical therapists are well enough equipped to apply the specific behavioural aspects of the BGA programme in daily practice. In this trial, the physical therapists were trained to provide an adequate BGA programme in a two-day training course, with three days of feedback during the trial period. The question that remains is whether this training course was sufficient to equip physical therapists well enough to provide a BGA programme. More specifically, we wonder whether the course succeed in changing the beliefs and attitudes of the physical therapists in such a short period of time. The attitudes and beliefs of the physical therapists were assessed according to the Pain Attitudes and Beliefs Scale for Physiotherapists (PABS PT)30, a questionnaire that was originally developed for health care providers treating patients with low back pain and which we adapted for neck pain. The questionnaire focuses on two aspects: biomedical treatment orientation and biopsychosocial treatment orientation. Although the trained physical therapists were more bio psychosocially orientated compared to the manual therapists, the differences were small. (table 1)
Table 1. Score of the Pain Attitude and Beliefs Questionnaire for physiotherapist (PABS PT) after specific training in the BGA protocol

<table>
<thead>
<tr>
<th>PABS PT score</th>
<th>Manual therapy</th>
<th>BGA programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>biomedical (10-60)</td>
<td>manual therapists (n=12)</td>
<td>physical therapists (n=8)</td>
</tr>
<tr>
<td>biopsychosocial (9-54)</td>
<td>28.4 (8.7)</td>
<td>17.4 (5.6)</td>
</tr>
<tr>
<td>mean (SD)</td>
<td>36.0 (6.4)</td>
<td>42.2 (4.6)</td>
</tr>
</tbody>
</table>

To answer the question of whether the therapists adhere to the strict BGA protocol, we performed a treatment-integrity check. The therapists filled in a treatment registration form for each treatment session, and the results showed that in 40.8% of all cases the BGA treatment was not provided adequately. The BGA programme can be divided into three phases; 1) the initial phase, which involves a reconceptualisation of the patient’s pain model, 2) the treatment phase, which can be characterised as a time-contingent increase in activities from baseline towards pre-determined goals and 3) the generalisation phase, which aims to encourage the patients to proceed with their healthy behaviour during the activities of daily living\textsuperscript{21}. Especially in the treatment phase, in which the patients tried to achieve their pre-determined goals, and during the generalisation phase the physical therapists seemed to have problems in adhering to the strict BGA protocol. Many therapist changed the BGA programme into a more biomedical, hands-on approach and some therapists even left out the generalisation phase. There are several reasons for the non-adherence by the therapists. One reason for not adhering to the BGA programme could be the quick recovery of the patients (60% of the patients were recovered after 6 weeks). Another reason, especially for physical therapists who graduated in physical therapy many years ago, could be that they were educated in the biomedical model, on which they have based their treatment for many years. Therefore, they might be set in their ways. But even the more recently graduated therapists were still trained to think according to the biomedical model, which influences their attitudes and beliefs. However, in the future this might change, because in physical therapy training more and more attention is currently being
Chapter 11

paid to the bio psychosocial model. Furthermore, another reason for not adhering to the BGA protocol could be the organisation within the physical therapy clinic, which does not always allows such a programme to take place, despite the substantial efforts made by all the physical therapists in this trial. After all, patients have to perform various activities, for which it is sometimes necessary to have sufficient room, or specific chairs, and desks with computers, etc. An additional problem can be the frequency of treatment sessions in the BGA programme. Since frequent reinforcement of behaviour is necessary, the BGA treatment must be frequent. A BGA programme must consist of enough therapy sessions to achieve not only treatment results and pre-determined goals, but also generalisation of behaviour\textsuperscript{43}.

A lack of time available for the therapy also seemed to be a problem. In most cases only 25-30 minutes were available for a treatment session, which is rather short to execute and establish all the components of a BGA programme. At present, in daily physical therapy practice it is difficult to change the duration of treatment session, so if therapists want to work according to the BGA programme they have to be aware of this problem.

As mentioned above, one of the reasons why the physical therapists deviated from the protocol was the quick recovery of the patients. This indicates that the pre-determined therapy goals were already achieved in a shorter time than was expected, and after less therapy sessions. It can therefore be concluded that either the intensity of the BGA programme was too low for the presented complaints, or the number of therapy sessions needed was over-estimated. A priori we aimed at 18 sessions in 9 weeks. According to the supervisors of the BGA training course and the psychologist, it takes this amount of sessions to change behaviour and to complete the generalisation phase. However, in daily practice in a primary care setting this strict planning of the number of sessions within a certain time-frame is difficult.

Patient related factors can also influence the BGA programme. Veenhof et al\textsuperscript{40} concluded on the basis of a qualitative study that initial motivation for a long-term result was associated not only with adherence to the programme but also with the involvement of patients during the treatment. Unfortunately, we did not measures this initial motivation in our study, so we can neither confirm nor refute this conclusion. However, although the study focussed on patients with chronic pain patients it seems likely that this factor can also play a role in patients with sub-acute neck pain.
We conclude that for strict adherence to the BGA programme additional training is necessary in order to equip physical therapists well enough to provide a BGA programme and to change their attitude into a more biopsychosocial attitude.

**Manual Therapy**

The manipulative components of the MT were the same as in the Hoving trial. However, what was different from the Hoving trial was the fact that the MT intervention in the present trial was, in nearly all cases, combined with exercises and advice. The exercises consisted mainly of stabilisation and mobilisation exercises and activities to be performed at home. The advice included: an explanation of the pain (pain is not dangerous), advice to stay active, and advice concerning work situation or sport activities. We also performed a treatment-integrity check on the MT treatment and the outcome was that this treatment did reflect usual care MT for neck disorders in the Netherlands, as described by Oostendorp, and according to the Manual Therapy Competence Profile, which is underpinned by the best available evidence.

In line with the findings of former studies, we consider MT to be effective for the treatment of patients with sub-acute neck pain. However, it seems remarkable that a therapy which includes passive mobilisation and manipulation techniques, mainly based on clinical expertise, is so effective. We would recommend more fundamental research to explore the `black box` of MT.

**Comparison of Manual therapy with the BGA programme**

In the present study, MT was less effective in the treatment of patients with sub-acute neck pain than BGA treatment, the differences were only significant on the outcome disability. It is possible that the small differences between the two interventions, namely the fact that both consisted of active exercises and included advice, is responsible for the small difference in effect. Furthermore, it can be questioned whether the remarkable results of these two interventions are due to spontaneous recovery from sub-acute neck pain. However, comparing the results with the results of the Hoving trial, the patients in the present trial showed nearly 20% more improvement in both groups. So, based on this comparison we assume that the natural course is not fully responsible for the results of the two interventions. A recommendation can therefore be made for future research, to obtain more insight into the influence of the natural course of sub-acute neck pain, a RCT including a no treatment group should be conducted.
General practitioners
The GPs were recruited by the principal investigator and 72 agreed to participate. After they had been contacted personally, the GPs agreed to ask all patients with neck pain who did fulfill the inclusion criteria for this trial to participate. Precautionary actions were undertaken to make the inclusion as easy as possible with very little extra work for the GP. The expectation was that it would be easy to include the number of patients needed for the trial. However, despite all the actions that were undertaken, it proved to be far more difficult, for several reasons. One of the reasons could be the change in health care legislation in the Netherlands that was made during the inclusion period. The new legislation de-listed physical therapy from basic health care insurance, and patients therefore had to have additional health insurance to cover the costs of physical therapy. This and the fact that many GPs did not immediately understand the implications of this new legislation system for their patients, may be a reason of the low inclusion rate. Although the GPs were informed about the new legislation in our frequent newsletters (n=22), the recruitment remained problematic during the whole trial period. A second reason could be that in their daily work GPs do not focus on the inclusion of patients for research purposes; 37 GPs (51%) did not include any patients at all. The number of patients included by the other 35 GPs (49%) varied from 1 to 20 patients per GP; a few GPs (3) succeeded in including more than 15 patients. The variation in the rate of inclusion among GPs was therefore enormous (Figure 1).

Figure 1. Number of patients included per general practitioner
Population
As mentioned before, taking into consideration the results of the Hoving trial, in which 48% of the included patients had sub-acute neck pain, and the fact that the transition from acute to more chronic pain is believed to take place within this time-frame\textsuperscript{10}, we decided to include patients with sub-acute neck pain. Gatchel\textsuperscript{10} describes a conceptual model in which emotional reactions, such as fear, anxiety and worry, are associated with the consequence of the perception of pain during the acute phase. This model has been evaluated and confirmed by other authors\textsuperscript{9;22;25;27;31;32;37;44}. Although most of the research on this topic has focussed on patients with low back pain, psychological factors may act as mediators in the development of chronic pain, and these are believed to be the same for low back pain as for neck pain. However, the question that arises is whether this assumption is correct. One of the main findings in the present trial was the low baseline score of the patients on most outcome measures related to this conceptual model. For instance, overall this primary care population demonstrated no fear of movement, distress, lack of coping, etc. Patients with neck pain and especially sub-acute neck pain in a primary care setting probably differ in their attribution and beliefs from patients suffering from low back pain. Sub-acute neck pain seems to be less disabling, and therefore these patients seem to be less vulnerable for chronicity. Relatively little is known about the specific mechanisms through which sub-acute neck pain will become chronic neck pain, or about the match of specific treatment programmes to patient characteristics\textsuperscript{45}. We hypothesise that in patients with high scores for psychological mediators such as fear of movement, distress, depression, etc, BGA programme can be the most promising strategy.

Patient expectations are also believed to be an important issue, and these expectations could influence a patient’s compliance with therapy. Expectations originate from several sources, among others from former experiences, information from the social environment (e.g. family or colleagues), or information from the media, and of course also from information provided by physical and manual therapists. It is logic that the information provided by the therapists and the GPs will be influenced by the therapist’s attitude. If a therapist states that he or she will “fix the problem”, this attitude and information is likely to have an effect on the patient’s expectation as well as the patient’s perception of the therapy. We therefore emphasise that it is important that all therapists are aware of their own attitudes and expectations and the possible effect that they may have on the attitudes and expectations of the patients.
Unfortunately, we did not ask the patients about their expectations in advance, but we recommend that patient expectation should be studied in future research.

We did ask the patients about their preferences for therapy before randomisation: 65.1% of the patients had no preference for either type of therapy, 12.3% had a preference for physical therapy and 22.6% had a preference for MT. The size of the sub-groups was too small to draw conclusions about the differences in outcome of the two therapies.

**Recommendations**

- Future research should test the effectiveness of a BGA programme in patients with fear of movement, distress, depression, or other psychological mediators, who might be more responsive to a BGA programme, than patients with a low score on these parameters.
- Patient preferences and expectations are likely to influence the outcome, so qualitative research on these issues would add new information to increase our understanding of this concept.
- In order to optimise their provision of a BGA programme therapists should specialise in this specific approach, and follow training courses in order to change their attitudes and beliefs about treating patients. Apart from the behavioural aspects, the cognitive aspects that are relevant for therapists should also be included in this specialist training.

**Prognostic factors**

The second aim of this thesis was to investigate which factors influence the short and long-term course of neck pain. The results of the study on prognostic variables adds to our understanding of how the course of the complaints can be influenced. In addition to the known factors, such as the history of the complaints or the severity of the complaints, psychological mediators are also expected to influence the transition from acute to chronic pain.

**Main findings**

- No core set of prognostic psychological factors was found to predict the outcome of neck pain after one year. Consequently, our study does not provide new information about which factors
are important in the transition from sub-acute to chronic neck pain. The only factor that was found to be of marginal importance was fear of movement (Chapter 5).

**Discussion**

The literature describes a diversity of factors that are considered to influence the course of neck complaints. As Linton stated in an extensive review, psychological factors are also related to neck pain and back pain from its inception to the chronic phase. Furthermore, psychological factors are thought to be pivotal in the transition from acute to chronic pain, as well as influential in the onset of pain. Although this is a seemingly logical statement, it is hard to find any consistent evidence in the literature. This can be explained by answers to the different research questions addressed in these studies. For example, if a study focussed on a working population, only work-related factors were involved. In the Hoving trial the participants were selected by the GPs, so clinical variables were mainly examined. In both of these above-mentioned studies, psychological mediators were not the primary objective, whereas in our trial special attention was paid to psychological mediators such as fear, distress, depression, somatisation, catastrophising, etc. In our study, the most important prognostic variables differed per outcome of interest, which is in line with the findings of other studies. This suggests that pain, disability and perceived recovery are all really different concepts. The explained variance of the prognostic models in the present study was low (16%-30%), probably due to the homogeneity of the study population and the low baseline score for most prognostic factors. Therefore, we were unable to confirm Linton’s, statement. It would be interesting to investigate whether psychological factors can influence the course of neck pain in a population with more variation in their scores for the psychological factors than to the population in the present study.

**Recommendation**

- Further prognostic research is needed to find more consistent answers. Taking into account the fact that all biopsychosocial variables have the potential to influence the course of the complaints, it seems to be logical to take all these variables into consideration, especially in a population with more variation in the potential prognostic factors.
Clinimetric properties

The third aim of this thesis was to assess the reproducibility and validity of frequently used clinical measurements and questionnaires, focusing on neck pain, neck function and related disabilities. This is of importance, since therapists base their treatment strategy on scores, or the results of clinical tests and questionnaires. However, probably due to the variation in patients and therapists it is difficult to obtain reliable tests results. Furthermore, the validity of most of the frequently used tests is unknown or poor. To decide which treatment strategy should be used, most therapists rely on physical examination, their clinical expertise, and the opinions of peers within their field of practice. Although clinical expertise is a structural part of evidence-based practice\textsuperscript{34}, valid and reliable instruments are important tools in terms of objectivity and transparency.

We will discuss the reliability of some frequently used tests, the interpretation of the patients’ answers on the questionnaires, and the interpretation of score changes on some of the questionnaires.

Main findings

- Assessment of general mobility and inter-segmental mobility in patients with neck has poor reproducibility, both in terms of agreement and reliability (Chapter 6).
- Measuring cervical range of motion in patients with neck pain with the EDI-320 inclinometer appears to have good intra- and inter-examiner reproducibility, in terms of agreement and reliability (Chapter 7).
- Qualitative research in the development and validation of questionnaires is an important way in which to enhance validity (Chapters 8 and 9).
- The estimated Minimal Clinical Important Change using the optimal cut-off point of the ROC curve is 3.5 for the Neck Disability Index (NDI) and 2.5 for the Numerical Rating Scale (NRS) for pain intensity (Chapter 10).
Discussion

As described in Chapter 6, there is poor agreement and reliability for the assessment of segmental mobility. This means that the conclusion of manual therapists after clinical assessment of a patient (for example, that there is a fixation, or a limitation in movement at a certain segmental level of the cervical spine) is doubtful. Furthermore, the validity of these segmental mobility tests is not known, due to lack of a reference standard. In the future, open, dynamic, Magnetic Resonance Imaging (MRI) may make it possible to perform validation studies and improve the performance of segmental mobility tests. Overall it can be concluded that the tests for inter-segmental mobility were not reproducible. Moreover, the results for the full range of motion were the same: the reproducibility was poor.

In this thesis, the reproducibility of a digital inclinometer, the EDI-320, was tested. This instrument assesses the mobility of the neck in terms of degrees in range of motion. Although this type of instrument cannot be used to measure segmental mobility, it measures full range of motion of the cervical spine instead. The EDI 320 inclinometer has a good inter- and intra-examiner reproducibility, as described in Chapter 7, and also previously described by Koes et al.20 However, it will be difficult to use this instrument in daily practice, because it is more time-consuming than the measurement of a cervical range of motion device (CROM). Most therapists will use the “best guess” method, a visual estimate, to assess the full range of motion, but unfortunately this measurement is insufficiently reproducible.

One reason for exploring limitations in movement or segmental movement is the causal relationship that is postulated between a assumed limitation and the complaints. The use of manipulations or specific mobilisations in daily practice is solely based on the results of segmental mobility tests, with poor reliability, and the expertise of the clinician, credited by the subjective perceived recovery of the patient. Consequently, as stated before, there is no evidence to support manipulations and mobilisations in the cervical spine. Taking this into account, high velocity thrust manipulations, especially in the higher regions of the neck, are not recommended, also considering the minimal risks involved.3;33 Mobilisation is based on skilled, low-grade passive movement, and is considered to involve less risks.

The clinimetric properties, i.e. the reproducibility and the validity of the NDI and the NRS, were found to be good6;13;42. Using these questionnaires, information can be obtained from patients with neck
pain about their perceived pain and disability. However, the ability of the NRS and NDI to detect changes over time, the minimal clinically important change (MCIC), has been a topic of debate. We used two methods to investigate the MCIC, integrating both an anchor-based and a distribution-based approach: the Minimal Detectable Change (MDC) and the optimal cut-off point of the Receiver Operator Characteristic (ROC) curve. For the anchor-based strategy, an anchor or external standard was defined to classify patients as stable or unchanged. We used the global perceived effect (GPE) score as an anchor. The GPE is widely used as an outcome measure, and appeals to the patient’s perception of recovery. However, it can be questioned whether the use of the GPE, which is a transition scale, to classify stable patients is correct. First of all, not only patients who experienced a slight recovery, but also patients who experienced a slight worsening of their complaints were classified as stable. This implies that within the stable group there was a distribution in the representation of complaints present. Although this distribution may have influenced the results of the study, a more appropriate method to define stable patients was not available. We could define only patients who scored no change at all as stable patients, but to define also slightly recovered patients as stable is a choice based on the assumption that patients will also give socially desirable answers, and will therefore be more inclined to answer ‘slightly recovered’ instead of ‘no change at all’. The GPE score can also be influenced by the difficulty some patients have in remembering their former health status; a recall of six months or more is rather a long period over which to compare the current health status with a former one.

A possible alternative for the use of the GPE score could be qualitative research in which a protocolised interview technique can be used to identify patients who consider themselves recovered or not. However, the disadvantage of qualitative research is that the process is time-consuming, and therefore costly. Although the same arguments apply to a recall f for example of six months, there is a greater possibility to obtain more information about the patient’s recovery.

Using an anchor-based strategy we found a high MDC value on both the NDI and the NRS questionnaires. This could be explained by the substantial measurement error for the unchanged patients, since all variations within patients, and within the therapists, including measurement errors, have been taken into account. Furthermore, most of the patients who recovered during the study demonstrated a change which lies within the measurement error of the questionnaire.
We chose to use the ROC curve, with both a distribution-based and an anchor-based strategy, which corresponds more with the clinical situation, because not only stable patients, but also patients who had recovered were taken into account. The mean NDI score at baseline was 14, on a scale which varies between 0-50, and is considered to reflect a mild disability\textsuperscript{41}. The MCIC, based on the ROC curve was 3.5 points. This indicates that a change in mean score from 14 to 11.5 on the NDI can be considered as an important change. However, it should be mentioned that, although a clinically important improvement occurred, the mean score on the NDI after the intervention still implied mild disablement. After 52 weeks the mean score on the NDI was 5.0 points. According to Vernon, this score still indicates mild disablement, although the mean score differed 9.0 points from the baseline measurement. The question that arises is whether the Vernon’s classification of the NDI can be used in the evaluation of patients.

Another issue was the validity of psychological questionnaires. Many patients in the present trial did not understand some items in the TSK and the PCCL questionnaires. The fact that many patients experienced these problems was the reason why we performed a qualitative study. After all, a different interpretation of the items by patients will have an effect on the validity of both questionnaires. A structured interview method, the Three Step Test Interview (TSTI) was used to test the interpretation of items on the TSK and the PCCL. The outcome changed our view with regard to the applicability of these questionnaires. The patient’s perception of words like “injury” and “dangerous” differed from that of the researchers. Others problems were: long complicated formulations, composite questions, irrelevant questions, no frame of reference, and wrongly interpreted questions.

Using questionnaires in a large population with a various different ethnic and cultural backgrounds and levels of education, has in our opinion influence on the outcome of the questionnaire. We conclude that qualitative methods have an added value in the development of self-report questionnaires. After all, certain problems can be identified that will be missed if only quantitative questionnaires are used. We recommend that a qualitative study should be fully integrated in the development of quantitative questionnaires.

Although quantitative researchers criticise on the sample size of qualitative research, because of doubts about the generalisability of the results, qualitative research is nevertheless worthwhile. It is a descriptive way of analysing and understanding problems that can arise
from differences in patient perceptions, using structured interview techniques.

**Recommendations**

- Valid and reproducible measurement instruments should be developed for the clinical assessment of patients.
- We recommend a mix of methods when developing self-reported questionnaires, because qualitative methods have an added value in that they highlight problems that cannot be identified through the sole use of quantitative methods.

**Clinical relevance of this thesis**

Although the amount of literature on neck complaints does not equal the amount of literature on low back pain, there is clearly a need for clinical guidelines for treatment of this problem in the Netherlands. Physical therapy guidelines on Whiplash Associated Disorders are already available, but patients with this complaints differ from those with regular neck disorders. However, guidelines for mechanical neck disorders are available in Australia. Based on the available literature, including the information contained in this thesis on clinical assessment, i.e. concerning segmental mobility, measurement instruments, questionnaires, and interventions, the development of multidisciplinary clinical guidelines on neck disorders comes within reach.

In this thesis, multiple assessment tools such as segmental mobility tests and questionnaires were tested for reproducibility, validity and interpretation of the results. This is relevant for all clinicians who assess mechanical neck complaints in daily practice. We advise these clinicians to use measurement tools with good clinimetric properties, but if they are still using tests with low reproducibility or non-evidence-based validity, they must be aware that no firm conclusions can be drawn from these tests. From this thesis we conclude that segmental mobility tests are not useful, because their reproducibility is minimal and their validity is unknown.

With regard to the choice of therapy, we found only a slight difference in effectiveness between MT and BGA treatment in favour of the latter. Patients who followed a BGA programme scored slightly better on the disability scale. However, in order to set up a BGA programme in primary care, the therapist must make a considerable investment in terms of a change in the organisation of the clinical setting, training, education and effort. To provide MT it is also necessary to train skills
and to gain extra knowledge, although in the Netherlands MT is included in usual care. In the current situation a GP might refer a patient who is suspected of fear of movement or other psychological problems to a physical therapist who is trained in BGA treatment, and other patients to a manual therapist.
Chapter 11

Reference List


204


Chapter 11


Summary
Summary

Is neck pain a “pain in the neck” for clinicians and therapists?
To answer this question the focus of this thesis is twofold. The first part
of this thesis concerns an overview of interventions and describes the
design and the results of a randomised clinical trial. The second part
concerns clinimetric properties of frequently used clinical tests and
questionnaires.

1. Introduction

In Chapter 1 an outline of topics covered in this thesis is given.
Furthermore, the research questions are described.
Considering neck pain, a common musculoskeletal disorder, there still is
a lack of evidence in favour of any treatment modality. Also with regards
to the quality of various frequently used clinical test and questionnaires
there are still many questions unanswered.
Sub-acute non-specific neck pain patients are an interesting subgroup,
since from 4 to 12 weeks a transition from acute to chronic neck
complaints takes place. It has been hypothesised that psychological and
social factors play an important role in this transition. With this in mind,
it is a challenge to study the effectiveness of an intervention which
focuses specifically on these psychological and social factors. In the
study, described in this thesis, we compare a behavioural graded activity
programme with manual therapy. The latter, despite the lack of strong
evidence, is the most effective therapy for neck pain so far.
The evaluation of the quality of clinical tests and the study of the
applicability of questionnaires used in the clinical setting is also an
objective of this thesis.

2. An update

In Chapter 2, we discuss the benefit of evidence based medicine in
general, and its role in the treatment of neck pain. Although much
evidence for conservative therapy for neck pain is inconclusive,
manipulative therapy and/or mobilization in combination with exercise
seem to have the most promising results. Additionally, manipulative
therapy would appear to be more cost-effective than physical therapy or
standard medical care (as administered by the general practitioner).
3. Study protocol

The design of a randomized clinical trial comparing a behavioral graded activity programme with manual therapy for patients with non-specific sub-acute neck pain is described in Chapter 3. The behavioral graded activity programme is a time-contingent approach with an increase in activities from baseline towards pre-determined goals. This protocol was developed in cooperation with experts in the field of behavioral medicine. Manual therapy consists mainly of specific spinal mobilisation techniques and exercises. Patients were included in the study when non-specific neck pain persisted more than 4 weeks but no longer then 12 weeks; aged between 18 to 65 years; when they were referred by the general practitioner.

Primary outcomes included global perceived effect, pain intensity and functional disability. Secondary outcomes included various psychological characteristics, such as kinesiophobia, somatisation and distress.

Patients completed questionnaires at baseline, 6, 12, 26 and 52 weeks after randomisation.

4. Effectiveness

In Chapter 4 the effectiveness of a behavioural graded activity (BGA) programme compared to manual therapy (MT) is described. 35 general practitioners recruited 146 patients with sub-acute non-specific neck pain. Patients were allocated to either the BGA group (n=71) or the MT group (n=75). Overall, the multilevel analysis showed a marginally but statistically significant difference on the outcome disability, mean difference for Neck Disability Index = 2.42 (95% CI: 0.52-4.32), and at 52 weeks, on the outcome pain, mean difference on the NRS = 0.99 (95% CI; 0.15-1.83), both differences are in favour of the BGA programme.

The analysis did not show a significant effect on all other outcomes. The success rates at 52 weeks, based on the GPE were 89.4% for the BGA programme and 86.5% for MT.

We encountered numerous practical problems such as a poor compliance of the physical therapists to perform the BGA protocol. Based on this trial it can be concluded that there are only marginal, but not clinically relevant, differences between a BGA programme and MT.
5. Prognostic factors

An increasing amount of attention is being paid to psychological factors in the development and maintenance of pain and disability. Furthermore, psychological and social factors are possibly involved in the transition from acute to chronic neck pain. In Chapter 5 a prospective study is described. The aim is to assess whether psychological factors are prognostic indicators of the short and long term outcome. 146 patients with non-specific sub-acute neck pain are included.

Multilevel analyses showed very diverse results on the outcomes perceived recovery, pain and functional disability. Only ‘fear of movement’ turned out to be consistently and significantly present in the univariable analysis for all outcomes at both follow-up measurements. Also for the short term In the multivariable analyses fear of movement turned out to be a prognostic factor for the outcomes pain and disability. We conclude that we were unable to identify a core set of prognostic psychological factors that predict the short and long term outcome of sub-acute neck pain. Further prognostic study are recommended.

6. Reproducibility

In Chapter 6 the inter-examiner reproducibility of physical examination of the cervical spine was assessed. Two physiotherapists independently judged the general mobility and the inter-segmental mobility (segments C0-T2) of the cervical spine. Furthermore, the provoked pain score of the patients during each test was recorded. Despite the use of a standardised protocol to assess general mobility and inter-segmental mobility of the cervical spine it is difficult to achieve reasonable agreement and reliability between two examiners. Agreement for general mobility shows kappa values between 0.05 and 0.61, and for the inter-segmental mobility it shows kappa values between –0.09 and 0.63. Likewise, the patients are not able to score the same level of provoked pain in two assessments with an interval of 15 minutes. The ICC’s varied between 0.36 and 0.71 for general mobility and between 0.22 and 0.80 for inter-segmental mobility.

We conclude that despite the use of a standardised protocol to assess general mobility and inter-segmental mobility of the cervical spine it is difficult to achieve reasonable agreement and reliability between two examiners.
7. Reproducibility of the range of motion

An assessment of the intra-rater and inter-rater reproducibility of the measurement of active Range of Motion (ROM) in patients with neck pain using the Cybex Electronic Digital Inclinometer-320 (EDI-320), is described in Chapter 7.

In an outpatient clinic in a primary care setting 32 patients with at least 2 weeks of pain and/or stiffness in the neck were randomly assessed. In a test-retest design with blinded raters a standardized measurement protocol was used to test cervical flexion-extension, lateral flexion and rotation.

In general, the intra-rater reproducibility and the inter-rater reproducibility were good. Reliability showed an Intraclass Correlation Coefficient of 0.93 or more for intra-rater reliability and 0.89 or more for inter-rater reliability. The 95% limits of agreement for intra-rater agreement, expressing the range of the differences between two ratings were -2.5 ± 11.1° for flexion-extension, -0.1 ± 10.4° for lateral flexion and -5.9 ± 13.5° for rotation. For inter-rater agreement the limits of agreement were 3.3 ± 17.0° for flexion-extension, 0.5 ± 17.0° for lateral flexion and -1.3 ± 24.6° for rotation. In general we conclude that the intra-rater reproducibility and the inter-rater reproducibility was good.

8. Qualitative research

During the performance of the randomised clinical trial problems were encountered by the patients filling in questionnaires. The aim of the study described in Chapter 8 is to elicit these problems. Patients understanding and interpretation of the wording used in test items of the Tampa Scale of Kinesiophobia (TSK) are evaluated using a qualitative method: the Three-Step Test Interview (TSTI), which is an observational technique that aims to discover problems with self reported questionnaires. The TSTI consists of three phases: 1) concurrent think aloud; 2) a retrospective interview; 3) a semi structured interview. Through the TSTI data are collected with regards to the thoughts or considerations of respondents while completing a questionnaire like the TSK.

The TSK was developed to measure fear of movement in patients suffering from low back pain. The TSK is being increasingly used for other pain conditions. In the analysis, each transcribed interview was divided into three segments. The thoughts and considerations were then analysed and categorised per item.
During the TSTI two problems were identified. One concerned the meaning of specific words used, like “dangerous” and “injury”. The other problem was that several implicit assumptions within some items make it difficult for respondents to answer these items.

It was concluded that in the development and validation of questionnaires like the TSK, not only quantitative psychometric properties are important, but also qualitative research has an important contribution to enhance applicability.

9. Qualitative research methodology

Psychometric analyses, such as factor analysis, internal consistency and construct validity, are well known and frequently applied methods in the development of health related patient reported outcomes. These statistical indexes shed hardly any light on how respondents interpret individual items, or on the meaning of their responses. In the study described in Chapter 9, the Pain Coping and Cognition List (PCCL), a quantitatively validated psychological questionnaire developed for chronic pain, has been subjected to a qualitative research method: the Three Step Test Interview (TSTI), similar as described in chapter 8. Six different types of problems were distinguished: long complicated formulations, composite questions, irrelevant questions, lacking frame of reference, problematic words, and wrongly interpreted questions. This study illustrates that quantitative methods have an added value when developing self reported questionnaires because problems were highlighted that can not be identified using quantitative methods only. Therefore, we recommend that a full qualitative study should be an integral part of the development of questionnaires.

10. Minimal Clinically Important Change (MCIC)

In Chapter 10 the minimal clinically important change (MCIC) on the Neck Disability Index (NDI) and the Numeric Rating Scale (NRS) for pain in patients with neck pain was assessed. Both measurement instruments are frequently used in research and clinical practice, but it is still unknown which changes are clinically relevant.

The MCIC was estimated with two different methods, both integrating an anchor-based and distribution-based approach: the minimal detectable change (MDC) and the optimal cut-off point of the ROC curve. The study population consisted of 183 patients with non-specific neck pain. The results show an MDC of 10,5 points for the NDI (scale
range 0-50) and 4.3 points for the NRS (scale range 0-10), and optimal cut-off points of the ROC curve of 3.5 for the NDI and 2.5 for the NRS. The estimated MCIC should be used as an indication for relevant changes in clinical practice. It was concluded that, using the optimal cut-off point of the ROC curve is a good choice, while false positives and false negatives are equally weighted.

11. General Discussion

In Chapter 11 the main findings of this thesis are summarized and reflected on. Challenges, practical and methodological issues of research are discussed in more detail. This chapter conclude with the clinical relevance for policymakers, physical and manual therapists as well as general practitioners, and recommendations for future research are made.
Samenvatting
Samenvatting

Is nekpijn een “pain in the neck” ofwel een groot probleem voor clini- en therapeuten?
Om deze vraag te beantwoorden is dit proefschrift in twee delen verdeeld. Het eerste deel geeft een overzicht over de effectiviteit van conservatieve behandelingen voor patiënten met nekklachten. Daarnaast wordt het design en de uitkomsten van een gerandomiseerd klinisch experiment beschreven.
Het tweede deel beschrijft onderzoek naar klinimetrische eigenschappen van veel gebruikte testen en vragenlijsten die betrekking hebben op nekklachten.

1. Inleiding

In hoofdstuk 1 worden de vraagstellingen die in het proefschrift aan de orde komen besproken. Daarnaast worden de begrippen nader toegelicht. Nekpijn is een veel voorkomend probleem aan het bewegingsapparaat. Er is nog weinig bekend over de werkzaamheid van behandelingen. Ook met betrekking tot de kwaliteit van vele vaak gebruikte diagnostische tests en vragenlijsten over nekklachten blijven vele vragen onbeantwoord.
Subacute niet specifieke nekklachten betreft klachten die bestaan van 4 tot 12 weken. Deze patiëntengroep is interessant omdat in die tijdsperiode ook de overgang van acute naar chronische klachten plaatsvindt. Als achterliggende reden van deze overgang wordt gedacht aan de invloed van psychologische en sociale factoren. Daarom lijkt het interessant om de effectiviteit van een behandeling die gericht is op de beïnvloeding van deze factoren te bestuderen. Om die reden wordt in dit proefschrift een gedragstherapie vorm van fysiotherapie, graded activity, vergeleken met manuele therapie, een therapie die tot nu toe als de meest effectieve therapie wordt beschreven. De evaluatie van klinische testen en de bestudering van de toepasbaarheid van een aantal veel gebruikte vragenlijsten is ook een van de doelstellingen van dit proefschrift.

2. De stand van zaken met betrekking tot de effectiviteit van behandelingen

In hoofdstuk 2 wordt de rol en voordelen van “evidence based medicine” besproken bij de behandeling van nekklachten. Er blijkt tegenstrijdig bewijs dat conservatieve behandelingen bij patiënten met nekklachten effectief zijn. Het meeste bewijs wordt gevonden voor een combinatie van manipulatie en/of mobilisatie technieken gegeven door de manueel
therapeut, indien deze gecombineerd worden met actieve oefentherapie. Daarnaast blijkt dat manuele therapie ook het meest kosteneffectief is indien het vergeleken wordt met een fysiotherapie behandeling of een behandeling uitgevoerd door de huisarts.

### 3. Onderzoeksprotocol

In Hoofdstuk 3 wordt het onderzoeksprotocol of design beschreven van het onderzoek naar de vergelijking van de effecten van een gedragsmatige graded activity programma met manuele therapie bij patiënten met sub-acute nekklachten. Het graded activity programma bestaat uit een tijd gebonden toename van activiteiten waarbij uitgegaan wordt van een basis-lijn meting en vooraf afgesproken therapie doelen. Dit protocol werd ontwikkeld in samenwerking met experts op het gebied van gedragstherapieën. Manuele therapie bestaat uit een combinatie van specifieke mobilisatie technieken aan de cervicale wervelkolom eventueel gecombineerd met actieve oefentherapie. Patiënten worden geincludeerd, indien ze niet specifieke nekklachten hebben, klachten welke niet korter dan 4 weken en niet langer dan 12 weken bestaan, een leeftijd hebben tussen de 18 en 70 jaar en indien ze verwezen worden door de huisarts. De primaire uitkomstenmaten zijn ervaren herstel, de pijnintensiteit en functionele status. Daarnaast worden een aantal uitkomstmaten meegegenomen die de verschillende psychologische karakteristieken, zoals bewegingsvrees, somatisatie en distress meet. Patiënten worden gevraagd vragenlijsten in te vullen op het moment van insluiting, en na 6, 12, 26 en 52 weken.

### 4. Effectiviteit

Hoofdstuk 4 beschrijft de effectiviteit van een graded activity programma vergeleken met manuele therapie. Door 35 huisartsen werden 146 patiënten met sub-acute nekklachten ingesloten in de studie. Patiënten werden na randomisatie toegewezen aan of een graded activity programma (n=71) of een manuele therapie behandeling (n=75). Na multilevel analyse bleek overall, een klein maar statistisch significant verschil te bestaan op de uitkomstmaat NDI (score van 0-50) in het voordeel van de graded activity behandeling, gemiddeld verschil op de NDI was 2.42 (95% CI: 0.52-4.32). Op de uitkomstmaat pijn (score van 0-10) bleek geen overall verschil te bestaan, wel op 52 weken bleek een statistisch significant verschil, gemiddeld verschil 0.99
Samenvatting

(95% CI: 0.15-1.83), opnieuw in het voordeel van het graded activity programma. Op alle andere uitkomstmaten bleken geen verschillen te bestaan. Op de uitkomstmaat “ervaren herstel” bleek na 52 weken 89.4% van de graded activity groep en 86.5% van de manuele therapie groep hersteld te zijn. Diverse praktische problemen zijn geïdentificeerd waaronder een beperkte therapietrouw door de fysiotherapeuten die het graded activity programma uitvoerden. Concluderend was de uitkomst van het onderzoek dat er marginale, maar geen klinisch relevante verschillen zijn tussen een graded activity programma en manuele therapie.

5. Prognostische factoren

Er is een toenemende belangstelling voor de invloed van psychologische factoren in de ontwikkeling van pijnklachten en beperkingen. Verder wordt aangegeven dat psychologische en sociale factoren een rol spelen bij de transitie van acute naar chronische klachten.

In hoofdstuk 5 wordt een prospectieve studie beschreven. Het doel van deze studie is te onderzoeken of psychologische factoren prognostische indicatoren zijn, ofwel het beloop van de klachten beïnvloeden zowel op de korte als lange termijn. 146 patiënten met sub-acute nekklachten werden geïncludeerd. Multilevel analyse geeft een verscheidenheid aan resultaten op de uitkomstmaten ‘ervaren herstel’, ‘pijn’ en ‘functionele status’. Alleen ‘bewegingsvrees’ blijkt een consistente en statistisch significante prognostische factor in de univariabele analyse voor alle uitkomstmaten op zowel de korte als de lange termijn. In de multivariabele analyses blijkt ‘bewegingsvrees’ alleen een prognostische factor op de uitkomstmaten pijn en functionele status, dit op de korte termijn.

De conclusie van de studie is dat er geen core set kan worden vastgesteld aan prognostische psychologische factoren voor sub-acute nekpijn. Dit geld zowel voor de korte als lange termijn. Verder onderzoek naar prognostische factoren wordt aanbevolen.

6. Reproduceerbaarheid

In hoofdstuk 6 wordt de reproduceerbaarheid van een fysiek onderzoek aan de nek tussen twee onderzoekers onderzocht. Twee fysiotherapeuten beoordeelden onafhankelijk van elkaar de algemene en segmentale (C0-T2) beweeglijkheid van de cervicale wervelkolom. Daarnaast werd de mate van pijn tijdens deze testen door de patiënt aangegeven en vastgelegd. Ondanks het gebruik van een
Samenvatting
gestandaardiseerd onderzoeksprotocol, bleek het heel moeilijk om een aanvaardbare overeenkomst tussen de onderzoekers in de bewegingstesten te demonstreren en om betrouwbaar te meten. De overeenkomst gemeten voor algemene mobiliteit gaven Kappa waarden die varieerden tussen de 0.05 en 0.61. De overeenkomst gemeten voor het segmentale onderzoek gaven Kappa waarden van -0.09 en 0.63. Ook de patiënten bleken niet in staat bij twee onderzoeken binnen 15 minuten eenzelfde pijnscore aan te geven. De Interclass Correlatie coëfficiënten (ICC) varieerden van 0.36 tot 0.71 bij de algemene mobiliteit testen en van 0.22 tot 0.80 bij de segmentale mobiliteit testen.

De conclusie van het onderzoek was dat het ondanks een gestandaardiseerd onderzoek protocol, van zowel de algemene als segmentale mobiliteit testen, het moeilijk is een redelijke overeenstemming te krijgen tussen twee onderzoekers, en betrouwbaarheid te meten.

7. Reproduceerbaarheid van de bewegingsomvang

In hoofdstuk 7 wordt de intra- en inter-tester reproduceerbaarheid van het meten van de beweeglijkheid van de nek beschreven. De beweeglijkheid is gemeten door gebruik te maken van de Cybex Electronic Digital Inclinometer-320 (EDI-320), een digitale hoekmeter. In een eerste lijn praktijk werden 32 patiënten met sub-acute neklklachten onderzocht door twee onderzoekers. Zoals beschreven in hoofdstuk 6 werd een test-hertest onderzoeksdesign gebruikt waarbij middels een gestandaardiseerd protocol de flexie/extensie, lateroflexies en rotaties van de nekwervelkolom werden gemeten. De intra- en inter-tester reproduceerbaarheid bleek goed waarbij een ICC van 0.93 voor de intra-tester reproduceerbaarheid en een ICC van 0.89 voor de inter-tester reproduceerbaarheid werd gehaald. De intra-tester reproduceerbaarheid, uitgedrukt in twee grenswaarden, was -2.5 ± 11.1° bij flexie-extensie, -0.1 ± 10.4° bij lateroflexie en -5.9 ± 13.5° bij rotatie. De verschillen tussen de twee onderzoekers, de inter-tester reproduceerbaarheid, was 3.3 ± 17.0° bij flexie-extensie, 0.5 ± 17.0° bij lateroflexie en -1.3 ± 24.6° bij rotatie.

De conclusie uit het onderzoek is dat er een goede intra- en inter-tester reproduceerbaarheid kan worden behaald indien gebruik wordt gemaakt van een EDI-320 om de beweeglijkheid van de nekwervelkolom te meten.
8. Kwalitatief onderzoek

Gedurende de uitvoering van de RCT bleken patiënten moeite te ondernemen bij het invullen van een aantal vragenlijsten. In hoofdstuk 8 wordt een kwalitatief onderzoek beschreven welke het doel heeft problemen die patiënten tijdens het invullen en interpreteren van vragen ondervinden in kaart te brengen. Het begrip en de interpretatie van woorden die gebruikt zijn in de vragenlijst naar bewegingsvrees (de TSK) worden in kaart gebracht. De TSK is ontwikkeld om bewegingsvrees te meten bij patiënten met lage rugklachten. De vragenlijst wordt echter steeds meer gebruikt bij patiënten met andere pijnklachten.

Er werd gebruik gemaakt van een kwalitatieve onderzoeksmethode de “drie-staps interview test” (TSTI). De TSTI bestaat uit drie fasen: 1) hardop denken gedurende het invullen; 2) een retrospectief interview over de antwoorden die gegeven zijn; 3) een semigestructureerd interview over de vragenlijst als geheel. De TSTI geeft op deze manier informatie over de gedachten en afwegingen van de patiënt die een dergelijke vragenlijst invult. Bij de analyse werden twee problemen geïdentificeerd; ten eerste de betekenis van woorden als “gevaarlijk” en “letsel”, en ten tweede de diverse aannames die worden gedaan in een aantal vragen waardoor het moeilijk was voor patiënten de vragen correct te beantwoorden. Geconcludeerd wordt dat bij de ontwikkeling en validering van vragenlijsten niet alleen kwantitatieve psychometrische facetten van belang zijn, maar dat ook kwalitatief onderzoek een belangrijke bijdrage kan leveren.
9. Kwalitatieve onderzoeks methodologie

Psychometrische analyses, zoals factor analyse, interne consistentie en construct validiteit zijn veel gebruikte methoden bij de ontwikkeling van gezondheid gerelateerde vragenlijsten als uitkomstmaat bij onderzoeken. Deze statistische analyses geven nauwelijks een indruk over de interpretatie van items in een vragenlijst of mening van de patiënt over sommige items. In Hoofdstuk 9 wordt de “pijn coping en cognitie lijst” (PCCL), een psychologische vragenlijst die kwantitatief gevalideerd is, onderzocht door gebruik te maken van de al eerder in hoofdstuk 8 beschreven TSTI methode. Zes problemen werden onderscheiden; lange gecompliceerde formuleringen, irrelevante vragen, gebrek aan een juiste referentie kader, moeilijke woorden en verkeerd geinterpreteerde vragen. De studie concludeert dat het gebruik van kwalitatief onderzoek een toegevoegde waarde heeft bij het ontwikkelen van vragenlijsten en derhalve wordt aanbevolen als methode naast de gebruikelijke kwantitatieve methoden.

10. Minimaal klinisch relevant verschil

In hoofdstuk 10 is het minimaal klinisch relevant verschil geschat van twee belangrijke uitkomstmaten bij patiënten met nekklachten; de Neck Disability Index (NDI) en de Numeric Rating Scale (NRS) voor pijn. Beide uitkomstmaten worden veel gebruikt zowel in onderzoek als ook in de klinische praktijk. Het is echter niet bekend welke veranderingen op beide schalen nu als een klinisch relevante verandering kan worden beschouwd. De klinisch relevante verandering is de kleinste verandering die patiënten als belangrijk ervaren. Twee methoden zijn gebruikt om de klinische relevante verandering te schatten; het minimaal te onderscheiden verschil (MDC) en het optimale afkappunt van de Receiver Operating Characteristic (ROC) grafiek. Gegevens van 183 patiënten met aspecifieke nekklachten werden gebruikt. De MDC voor de NDI blijkt 10.5 punten, dit betekend dat er 10.5 punten verschil moet worden gemeten op de NDI, een schaal die varieert van 0 tot 50 punten, wil men kunnen zeggen dat een patiënt zich verbeterd voelt. Voor de NRS pijn-schaal, welke varieert van 0-10, blijkt de MDC 4.3 punten. Indien het optimale afkappunt van de ROC-grafiek wordt gebruikt blijken deze cijfers 3.5 voor de NDI en 2.5 voor de NRS.
Samenvatting

Op basis van deze twee methoden wordt de conclusie getrokken dat het minimaal klinisch relevante verschil een indicatie kan zijn voor een gemeten verschil in de dagelijkse praktijk. Omdat er bij de methode waarbij het optimale afkappunt van de ROC grafiek gebruikt wordt gemaakt van zowel herstelde als niet herstelde patiënten, lijkt deze methode een goede keus.

11. Algemene discussie

In hoofdstuk 11 worden de belangrijkste bevindingen van het proefschrift samengevat. Daarnaast worden uitdagingen, praktische- en onderzoekstechnische aspecten van de onderzoeken bediscussieerd. Tevens worden de implicaties van de onderzoeken voor fysiotherapeuten, verwijzers, beleidsmakers en patiënten besproken en aanbevelingen gedaan voor toekomstig onderzoek.
Dankwoord
Het schrijven van een proefschrift is zoals bekend niet mogelijk zonder de hulp en steun van velen.

Het begon als lid van een projectgroep, in 1992, waar ik als bestuurslid van de Nederlandse Vereniging voor Manuele Therapie aan mocht schuiven om aan een onderzoeksvoorstel naar de effectiviteit van behandelingen bij patiënten met nekklachten mee te schrijven.

Hooggeleerde Koes, beste Bart, en Hooggeleerde Bouter, beste Lex, jullie gaven me toen al de indruk dat onderzoek doen toch wel erg leuk was en daarnaast een “must” voor de beroepsgroep om de effectiviteit van manuele therapie te onderbouwen. Met het onderzoek van Jan Hoving hebben we daar hard aan gewerkt en ik heb van jullie de mogelijkheid gekregen om gedurende dit traject als onderzoek assistent mijn steentje bij te dragen, maar ook om de POE cursus te volgen, wat voor mijn ontwikkeling van zeer groot belang is geweest. Dank voor het vertrouwen en de steun.

Hooggeleerde de Vet, beste Riekie, je hebt het toch maar aangedurfd een WOAiO, wat oudere AiO, aan te nemen om het vervolgtraject van Jan Hoving te doen. Je zult van mijn schrijftalent wel eens kramp hebben gehad, weer spaties vergeten of punten, haakjes in een tabel noem maar op, opgewekt kreeg ik altijd alles weer retour en dat in een hoog tempo. De kwaliteit die je eist is hoog en ik heb er erg veel van geleerd, ook wel eens gemopperd overigens, maar de sfeer was en is positief, werkend naar een goed product. De vrijheid die je me gaf was een verademing en daardoor was ik in staat alle nevenfuncties te continueren. Het ging om het eindproduct en dat hebben we mooi op tijd af gekregen. Dank voor alles, je steun, je kritiek, je opbeurende woorden en de structuur die je me gaf.

Beste Raymond, nog net niet gepromoveerd en dan al zo’n lastpost toegewezen krijgen om te begeleiden is niet mis. Ik ben je eerste promovendus en ik neem aan dat er voor jou nog vele volgen. Laat ik eens een opsomming maken (zoals je zo graag ziet), 1) ik bewonder je gevoel voor structuur, 2) je gevoel voor humor, 3) je werklust en 4) je altijd positieve inbreng. Heel erg bedankt daarvoor en ik wens je alle goeds toe met je gezin en in het vervolg van je wetenschappelijke carrière.
Dear Ann. We build up a dear friendship during all our meetings and congresses in the last two decades. I am grateful and proud that you are present during the defense of my thesis and I hope that in the future our professional and personal contact will continue. It is delightful to discuss matters on evidence based practice, on educational systems and on more continues collaboration between professionals throughout the world. Your contribution on those matters is enormous. But it is also delightful to exchange English and Dutch sayings and make silly translations. Our friendship is very special for me and Annelies, “I would not stick this under couches and chairs”. Thanks for everything.

Hooggeleerde Oostendorp, beste Rob, onze ontmoetingen stammen van vele jaren, in allerlei commissies en tijdens vele projecten. Jouw inbreng voor de manuele therapie is enorm geweest. Ik ben er trots op dat onze samenwerking daar af en toe aan heeft mogen bijdragen. Eén moment is onvergetelijk, ons bezoek met Ank en Annelies aan een weeshuis voor aids kinderen in een township in Kaapstad. Zo’n moment deel je maar met weinig mensen. Dank voor je samenwerking.

Hooggeleerde Vlaeyen, beste Johan, een paradigma shift werd het wel eens genoemd, maar dank zij jouw visie en ideeën heb ik toch een heel andere en veel genuanceerder kijk gekregen op chronisch pijn en pijngedrag. Je weet op een uiterst vriendelijk innemende manier altijd duidelijk te maken wat je bedoeld. Dank voor je begeleiding op afstand.

Hooggeleerde Dekker, beste Joost, ook wij kennen elkaar al lang, ik ben als broekie begonnen in de projectgroep van Margriet van Baar op het Nivel. Toen jij op de VU begon als hoogleraar zijn we elkaar weer vele keren tijdens de BOB tegengekomen. Dank voor je inzet en het doorworstelen van mijn manuscript.

Wieneke, Lilian, Kimi, Ilse, Geeske, Caroline, Martine, Judith en Sidney; lieve collega’s, wat is het fijn om in een inspirerende omgeving te mogen werken, waar altijd een helpinge hand was voor een syntaxie, een vraag of gewoon een lekker bakje koffie. Dank voor alle tijd, de luisterende oren, de fitness of de lunches. Wien, wij zijn 5 jaar geleden enthousiast begonnen, je bent er even tussenuit geweest maar volgend jaar is het jouw tijd, ik reken op je.

Lieve Nicole, ook jij was een erg inspirerende collega, doordat we een gemeenschappelijke interventie hadden in onze onderzoeken hebben we
Dankwoord

veel samen mogen doen en daardoor zijn we goede vrienden geworden. Leuk was het om te merken dat zo’n chaooot als ik toch heel goed met een gestructureerde en precieze collega kon samenwerken, we vulden elkaar goed aan. Je hebt me erg geholpen vooral met gewoon een luisterend oor. Ik ben er trots op dat we elkaars paranimf zijn.

Collega’s van IMPACT, een praktijk die in korte tijd twee promovendi produceert is een unicum. Dit kan alleen als men bereid is een hele hoop door de vingers te zien en achter de schermen steeds weer oplossingen te bedenken om de afwezigheid te compenseren. Ik waardeer dat echt enorm en hoop dat mijn inbreng de komende jaren voor de praktijk een stimulans zal betekenen.

Collega’s van de SOMT, als “Bijlist” te werken in wat vroeger nog wel eens als het vijandelijke kamp werd beschouwd, is heel apart. De werkomgeving van onze master opleiding is echt heel inspirerend en positief en we werken samen nu al twee jaar aan een degelijke en hoogwaardige master opleiding. Dank voor jullie vertrouwen en inbreng.

Ook zonder vrienden kan zo’n traject niet. Raymond en Ilse nu zijn we alle vier gepromoveerd, ik dan als laatste maar wat hebben we de afgelopen jaren gedurende onze vele etentjes geboomd, getwijfeld en gelachen. Heel erg bedankt voor jullie vriendschap. Ton en Patty, Ries en Claudia, Leo en Diane, Mike en Janine, Helma en Bert, Joop en Aafke, Frans en Tonny, Kees en Tiny, het is ook wel leuker om het eens niet over het vak te hebben, avondjes met een goed glas wijn, etentjes, weekendjes en dagjes uit zijn belangrijk geweest. Ook de enorme hulp die we van jullie hebben gehad bij het opbouwen van ons huis en de verhuizing zal ik niet gauw vergeten. Han, Jacq en Bert af en toe trainen voor de “ride of the roses” was heerlijk en ik hoop dat we nog vele fietsavondjes zullen krijgen.

Dank ook aan mijn ouders en schoonouders, zonder jullie was het niet mogelijk geweest. Bedankt Pa en Ma dat jullie me de mogelijkheid hebben gegeven te studeren en bedankt Pa Goudzwaard dat je ook bij mij in de commissie wilde plaatsnemen, ik vind het bijzonder.

Lisanne en Jesper, wat moet je nu met twee ouders die zo ambitieus zijn dat ze aan een promotie beginnen?? Jullie vermogen om je aan te passen aan een werkelijk vreemde en wisselende weekinvulling is gewoon geweldig. Gelukkig hadden we regelmatig op de boot onze tijd samen voor klaverjassen en andere spelletjes. Voordeel was wel dat we
Dankwoord

hele mooie reizen hebben gemaakt. Ik hoop dat er nog vele komen. Het
is fijn om zulke lieve prachtige kinderen te hebben.

Sabrina, Alex en Stephanie, tja, een vader die niet altijd even veel tijd
heeft om een bakkie te doen of even langs te komen is niet altijd even
gemakkelijk. Toch is ons contact altijd goed en positief gebleven. Ik ben
erg trots op wat jullie hebben bereikt. Bedankt voor alle hulp en begrip
zeker in het laatste hectische jaar.

Lieve Lies, we zijn nu al 20 jaar samen en wat is er veel gebeurd en
hebben we veel meegemaakt. Het spreekwoord “door dik en dun” is op
ons van toepassing, een onvoorwaardelijke liefde voor elkaar. Ik kon
altijd op je rekenen en ook de lay-out van dit boekje zou zonder jouw
hulp een ramp zijn geweest. Na jouw promotie zou het rustiger worden
voor je, je weet het resultaat dus dat zal ik niet gaan beloven na dit
traject. Maar ik hoop dat we nog vele jaren samen gelukkig kunnen zijn.
Dank gewoon voor alles.........
List of publications
List of Publications


Dutch publications


List of Publications


Curriculum Vitae
Curriculum Vitae

Jan Pool was born on December 14th 1954 in Voorburg, The Netherlands. In 1974 he studied Physiotherapy at the "Haagse Academie voor Lichamelijke Opvoeding". After graduation he worked in a private clinic in Zoetermeer. In 1980 he started the study Manual Therapy at the "School voor Manuele Therapie" in Utrecht. After graduation, he started lecturing Manual Therapy in Utrecht. He was a member of the board of the Dutch Association of Manual therapy in The Netherlands (NVMT), from 1990 till 1998. In 1997 he became a member of the Standard Committee of the International Federation of Manipulative Therapy (IFOMT). In 1995 he was involved as a member of the project group at the EMGO institute VU University medical centre in Amsterdam, starting a randomized clinical trial for the effectiveness of three interventions in patients with neck pain. He was involved in this study as a research assistant. He started the Postgraduate Epidemiology Programme at the EMGO institute and graduated as a master in epidemiology. In 2002 he started working as a PhD student at the EMGO institute. This resulted in several research articles in the field of musculoskeletal disorders including the articles covered in this thesis.

Currently, he is working in "Medical Centre Impact" in Zoetermeer as a manual therapist. Furthermore, he is a lecturer at the Master Education for Manual Therapy at the SOMT in Amersfoort and at the research centre Cobra in Amersfoort.

Jan is married with Annelies Pool-Goudzwaard and has four children, Sabrina, Stephanie, Lisanne en Jesper.