Neck-specific exercise with or without a behavioural approach, or prescription of physical activity in chronic whiplash associated disorders

Maria Landén Ludvigsson

Division of Physiotherapy
Department of Medical and Health Sciences
Linköping University, Sweden
Linköping 2016
To my family, Konny, Daniel and Emma,
and my parents Georg and Margareta

The sailor doesn’t ask for tail-wind, he learns how to sail
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>LIST OF PAPERS</td>
<td>3</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>4</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>5</td>
</tr>
<tr>
<td>Whiplash Associated Disorders (WAD)</td>
<td>5</td>
</tr>
<tr>
<td>- Classification of WAD</td>
<td>5</td>
</tr>
<tr>
<td>- Etiology and genesis</td>
<td>7</td>
</tr>
<tr>
<td>- Alterations of neck muscle behaviour</td>
<td>8</td>
</tr>
<tr>
<td>- Chronic WAD</td>
<td>10</td>
</tr>
<tr>
<td>- Treatment of WAD</td>
<td>10</td>
</tr>
<tr>
<td>Exercise and physical activity</td>
<td>11</td>
</tr>
<tr>
<td>- Neck-specific exercise</td>
<td>11</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>13</td>
</tr>
<tr>
<td>Outcomes</td>
<td>14</td>
</tr>
<tr>
<td>- Functioning and Disability</td>
<td>14</td>
</tr>
<tr>
<td>- Pain</td>
<td>14</td>
</tr>
<tr>
<td>- Psychosocial factors, focusing on self-efficacy</td>
<td>15</td>
</tr>
<tr>
<td>RATIONALE OF THE THESIS</td>
<td>17</td>
</tr>
<tr>
<td>AIMS</td>
<td>18</td>
</tr>
<tr>
<td>- General aim and hypothesis</td>
<td>18</td>
</tr>
<tr>
<td>- Specific aims</td>
<td>18</td>
</tr>
<tr>
<td>MATERIAL AND METHODS</td>
<td>19</td>
</tr>
<tr>
<td>Design</td>
<td>19</td>
</tr>
<tr>
<td>Participants, recruitment and randomization</td>
<td>19</td>
</tr>
<tr>
<td>Interventions</td>
<td>23</td>
</tr>
<tr>
<td>- Physiotherapist-guided neck specific exercise (NSE)</td>
<td>23</td>
</tr>
<tr>
<td>- Physiotherapist-guided neck-specific exercise with a behavioural approach (NSEB)</td>
<td>24</td>
</tr>
<tr>
<td>- Prescription of Physical Activity (PPA)</td>
<td>26</td>
</tr>
<tr>
<td>Data collection</td>
<td>27</td>
</tr>
<tr>
<td>Outcomes</td>
<td>28</td>
</tr>
<tr>
<td>- Main Outcome</td>
<td>28</td>
</tr>
<tr>
<td>- Secondary Outcomes</td>
<td>29</td>
</tr>
<tr>
<td>- Clinical relevance of outcome measurements and adherence</td>
<td>31</td>
</tr>
<tr>
<td>CONTENTS</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Other measurements – potential factors associated with outcomes</td>
<td>32</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>35</td>
</tr>
<tr>
<td>Statistical analyses</td>
<td>37</td>
</tr>
<tr>
<td>Ethical considerations</td>
<td>39</td>
</tr>
<tr>
<td>RESULTS</td>
<td>40</td>
</tr>
<tr>
<td>Disability and functioning</td>
<td>40</td>
</tr>
<tr>
<td>Pain</td>
<td>43</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>43</td>
</tr>
<tr>
<td>Factors associated with disability and pain reduction</td>
<td>45</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>48</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>51</td>
</tr>
<tr>
<td>Results – main findings</td>
<td>51</td>
</tr>
<tr>
<td>Disability/functioning, pain and self-efficacy</td>
<td>51</td>
</tr>
<tr>
<td>Factors associated with disability and pain reduction</td>
<td>53</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>55</td>
</tr>
<tr>
<td>Methodological considerations</td>
<td>56</td>
</tr>
<tr>
<td>Study Design</td>
<td>56</td>
</tr>
<tr>
<td>Classification of WAD</td>
<td>57</td>
</tr>
<tr>
<td>Clinical relevance and adherence</td>
<td>58</td>
</tr>
<tr>
<td>Analyses of those who declined participation and drop-outs</td>
<td>59</td>
</tr>
<tr>
<td>Outcome measurements</td>
<td>61</td>
</tr>
<tr>
<td>Cost-effectiveness analyses</td>
<td>62</td>
</tr>
<tr>
<td>Statistics</td>
<td>63</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>65</td>
</tr>
<tr>
<td>CLINICAL IMPLICATIONS</td>
<td>66</td>
</tr>
<tr>
<td>FUTURE RESEARCH</td>
<td>67</td>
</tr>
<tr>
<td>SUMMARY IN SWEDISH/SAMMANFATTNING PÅ SVENSKA</td>
<td>68</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>71</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>74</td>
</tr>
</tbody>
</table>
ABSTRACT

Background: Although 50% of those who suffer a whiplash injury still report neck pain after one year, there is a lack of knowledge about effective treatment for chronic whiplash associated disorders (WAD). Exercise is potentially useful, but the response to exercise in chronic WAD is highly variable between individuals and factors associated with good outcomes as well as the cost-effectiveness are unknown.

Aim: The general aim of this thesis was to evaluate the effect on self-reported disability/functioning, pain and self-efficacy of three different exercise interventions in chronic WAD grade 2 and 3, and to determine the cost-effectiveness of these interventions.

Material and methods: A total of 216 participants with chronic WAD took part in this randomized, assessor blinded, controlled, clinical trial. Participants were randomized to either neck-specific exercise without (NSE), or with a behavioural approach (NSEB), or prescription of physical activity (PPA) for 12 weeks. Evaluations of change scores and proportion of clinically relevantly improved participants regarding disability/functioning (Neck Disability Index (NDI)/Patient Specific Functional Scale (PSFS)), pain (Visual Analogue Scale of current neck pain (VAS-P), pain bothersomeness (VAS-B)) and Self-efficacy (Self-Efficacy Scale (SES)) were made after 3, 6, 12 and 24 months. Secondary analyses were made, regarding factors associated with clinically relevant improvements in disability, pain and regarding cost-effectiveness.

Results: Disability was more improved in the NSE/NSEB groups (NDI, P=0.02) than the PPA group, which reported no improvement, at 3 and 6 months, results remaining at 12 and 24 months (p ≤ 0.02)). Functioning (PSFS) was more improved in the NSE/NSEB groups than the PPA group at 3 months, in the NSEB compared to the PPA group at 6 months, and the NSE compared to the PPA groups at 12 and 24 months. The proportion of participants reaching clinically relevant improvement regarding NDI and PSFS was also larger in the NSE/NSEB groups at all time points (P<0.05), except NDI at 3 months and PSFS at 24 months. There were no differences between groups in VAS-P, VAS-B or SES change scores. The proportion of participants with clinically relevant reduction in VAS-P and VAS-B was however higher (P<0.02) in the NSE/NSEB groups compared with the PPA group at 3 and 12 months. Self-efficacy was only improved in the NSE group but without any between-group differences. There were no significant differences in any outcomes between the NSE/NSEB groups.
The only significant factor associated with both clinically relevant improvements in disability and neck pain both at 3 and 12 months was participation in the NSE group, with odds up to 5.3 times higher than in the PPA group. Different baseline features were associated with the improvements depending on the outcome and time point examined. From a societal perspective, NSE was the cost-effective option.

**Conclusion:** Physiotherapist-led neck-specific exercise resulted in better outcomes than prescription of physical activity regarding disability, functioning, and pain. The observed benefits of adding a behavioural approach to neck-specific exercise were inconclusive, and NSE was the cost-effective option from a societal perspective. Factors associated with clinically relevant improvements after exercise interventions in chronic WAD differed whether disability or neck pain was the outcome, but also differed in the short and long term. Participation in the NSE group was the only factor associated with both outcomes after both 3 and 12 months.


# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMG</td>
<td>Electromyography</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>EuroQol 5 Dimensions, Health related quality of life</td>
</tr>
<tr>
<td>HRQoL</td>
<td>Health Related Quality of Life</td>
</tr>
<tr>
<td>ICER</td>
<td>Incremental Cost Effectiveness Ratio</td>
</tr>
<tr>
<td>IMMPACT</td>
<td>The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials</td>
</tr>
<tr>
<td>MDC</td>
<td>Minimum Detectable Change</td>
</tr>
<tr>
<td>MCID</td>
<td>Minimal Clinically Important Difference</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NDI</td>
<td>Neck Disability Index</td>
</tr>
<tr>
<td>NSE</td>
<td>Neck-Specific Exercise group</td>
</tr>
<tr>
<td>NSEB</td>
<td>Neck-Specific Exercise with a Behavioural approach group</td>
</tr>
<tr>
<td>PCS</td>
<td>Pain Catastrophizing Scale</td>
</tr>
<tr>
<td>PPA</td>
<td>Prescription of Physical Activity group</td>
</tr>
<tr>
<td>PSFS</td>
<td>Patient Specific Functional Scale</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality Adjusted Life Year</td>
</tr>
<tr>
<td>QTF</td>
<td>Quebec Task Force on Whiplash Associated Disorders</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>SES</td>
<td>Self-Efficacy Scale</td>
</tr>
<tr>
<td>SF-6D</td>
<td>Short-Form 6 Dimensions, Health related quality of life</td>
</tr>
<tr>
<td>TSK</td>
<td>Tampa Scale of Kinesiophobia</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analogue Scale</td>
</tr>
<tr>
<td>VAS B</td>
<td>Pain bothersomeness, on a Visual Analogue Scale</td>
</tr>
<tr>
<td>VAS P</td>
<td>Current neck pain on a Visual Analogue Scale</td>
</tr>
<tr>
<td>WAD</td>
<td>Whiplash Associated Disorders</td>
</tr>
<tr>
<td>WAD grade 2</td>
<td>Neck pain and local physical neck findings</td>
</tr>
<tr>
<td>WAD grade 3</td>
<td>As WAD grade 2 + neurological deficits</td>
</tr>
</tbody>
</table>
Whiplash Associated Disorders (WAD)

One of the causes of neck pain, with a high risk of chronicity, is whiplash trauma. Over the past 30 years, the incidence of whiplash injury has increased in western countries and is estimated to be at least 300 per 100,000 in the adult population (1, 2). In Sweden the incidence of whiplash trauma is about 30,000 cases a year (3). It gives a broad effect on both society and for the individual in terms of suffering, health, productivity, and costs (1). About 50% of those who suffer a whiplash trauma continue to report neck pain one year after their injury (4), and among those with persistent symptoms about 50% experience persistent reductions in earnings relative to the trauma (5). Yet there is no clear evidence for any effective treatments for either acute, subacute or chronic Whiplash Associated Disorders (WAD), mainly due to lack of high quality studies (6, 7). The diagnosis and classification are challenging since there is no single test that verifies or excludes a whiplash injury.

Classification of WAD

The earliest description of something similar to a whiplash injury was the concept of “Railway spine” that was used to describe injuries conceived in train accidents in the 19th century. The actual concept of “whiplash” injury was first introduced in 1928 to describe cervical injuries that appeared after exposure of acceleration-deceleration forces in motor vehicle accidents (8). A modern definition, which by far is the most common classification system used in WAD research, and also commonly used in clinical practice, was stated in 1995 by the Quebec Task Force on Whiplash Associated Disorders (QTF) (9). This group consisted of a panel of experts in medicine, epidemiology, biostatistics etc. and they defined whiplash as:

“An acceleration-deceleration mechanism of energy transfer to the neck. It may result from rear-end or side-impact motor vehicle collisions, but it can also occur during diving or other activities or mishaps. The impact may result in bony or soft-tissue injuries (whiplash injury), which in turn may lead to a variety of clinical manifestations called whiplash associated disorders (WAD).”
The QTF can be seen as the golden standard classification system, and it classifies WAD according to type and severity of signs and symptoms observed in 5 different grades. It is also the classification used in this thesis:

- **0** No complaint about neck pain, no physical signs
- **1** Neck complaint of pain, stiffness or tenderness only, no physical signs
- **2** Neck complaint, musculoskeletal signs including, decreased range of movement and point tenderness
- **3** Neck complaint, musculoskeletal signs, neurological signs including decreased or absent deep tendon reflexes, muscle weakness, sensory deficits
- **4** Neck complaint and fracture or dislocation

In all grades other associated symptoms such as dizziness, headache, temporomandibular joint pain, tinnitus, deafness, memory loss and dysphagia can be found.

In WAD grade 2, local pain symptoms such as pain on palpation are required. There are many structures that can have been injured and cause neck pain (10). The majority of people with WAD present pain on manual palpation of various muscles of the neck. In contrast to those with chronic insidious onset neck pain or fibromyalgia, pain on palpation however appear more often in the upper part of the cervical spine in WAD (11).

In WAD grade 3 also neurological signs from the cervical spine are mandatory (9). The diagnostic criteria for neurological signs from the cervical spine or cervical radiculopathy, are not well defined however, and remain primarily a clinical diagnosis. Radiating pain is often part of the symptoms, and yet it is not uncommon that motor and sensory deficits without pain are the only symptoms (12). Testing of neurological signs is done clinically in every-day practice all over the world, but all clinical tests used are not proven to be either completely valid or reliable (13). Clinical findings are generally not very precise considering the substantial overlap of cervical nerve roots (14). Wainner et al. tested a number of tests for cervical radiculopathy in 82 participants with EMG-verified neurological impact. The tests included both traditional sensory tests, muscle testing, and muscle tendon reflexes, but also provocative tests such as distraction test, Spurling’s test of compression, and upper limb tension tests. Most individual items of the clinical examination were found to have at least a fair level of reliability, and several were found to have an acceptable level of accuracy.
However a cluster of tests is recommended, including Spurling’s test, upper limb tension test A, distraction test and range of movement (13).

Those with persistent symptoms following whiplash injuries form a heterogeneous group with variable and sometimes complex patterns of co-existing physical and psychological impairments (15, 16), and other classification systems have been suggested since the QTF system does not take psychosocial/behavioural aspects into account. For instance Sterling (17) and Poorbaugh et al (18) have suggested systems based on both psychosocial/behavioural and physical aspects while Söderlund and Denison merely suggest psychosocial/behavioural aspects in their classification system (19). These classifications are however not widely used, not even by those who proposed them.

**Etiology and genesis**

The WAD diagnosis is based on history taking and clinical examination. Even though there are symptoms and clinical findings that are more often associated with WAD than other neck disorders there is no single test that verifies the presence of a whiplash injury. The history of a sudden incidence that causes acceleration – deceleration forces to act on the neck is vital for the diagnosis, but since this often only involves a subjective description from the patient, the diagnosis has been questioned by some. J Dalton, the head of the Association of British Insurers, Motor and Liability department, claims that “The fact that whiplash is virtually impossible to disprove means that for too many it has become the fraud of choice” (20). However it is not even clear that a relationship between compensation-related factors and health in WAD truly exists. Studies show conflicting results and rarely consider reverse causality bias. Although it is commonly believed that claiming compensation leads to worse recovery, the opposite, that poor recovery may lead to compensation claims, is just as possible (21). Longitudinal data on neck pain up to 24 months post injury show that removing the financial incentive to over-report symptoms has no effect on self-reported neck pain in WAD (22), and compensation-seeking behaviour is not the main explanation for this group (5). Furthermore, there is a growing body of evidence of objective findings in people with WAD (23-30). Multiple anatomical sites in the neck have been postulated for a whiplash injury, including for instance neck muscles, facet joints, spinal ligaments, intervertebral discs, vertebral arteries and dorsal root ganglia (10). Abnormal increased rotational and translational segmental motions, verified by X-ray, have been found in women with WAD.
grade 1 or 2 compared to women without neck pain (26). One factor in particular, often attributed to the persistence of symptoms in individuals with WAD, is deconditioned neck muscles, which may affect the physical support of the cervical vertebral column (31). Higher neck muscle strength has been reported to be of importance to prevent neck pain in fighter pilots, who are exposed to high loads to the neck whilst flying (32). About 20-25% of the weight of the head is suggested to be stabilized by cervical ligaments, whereas muscles account for the rest (33).

**Alterations of neck muscle behaviour**

An important task of spinal muscles is to maintain the vertebrae within the neutral zone in which loading is optimally distributed to all supporting structure (33). A flattening of the normal lordotic curvature of the cervical spine is one of the most common radiographic findings after a whiplash injury (34), and this may affect the biomechanics of the muscles. The neck muscles are organized in several layers with different functions. The most superficial muscles of the dorsal side are the Trapezius and Levator Scapulae, which have attachments to the cranium and the cervical spine (Trapezius all levels, Levator Scapulae C1-C4). They contribute to neck movement, but are primarily considered muscles of the shoulder girdle (35), and have longer lever arms to the centre of movement in the cervical spine. The deepest muscle directly attaching to the cervical spine and thus with a short lever arm is the Multifidus. The Rotatores form the Transversospinalis muscle group together with the Semispinalis Cervicis and the Multifidus. They produce extension, ipsilateral side-bending and contralateral rotation of the neck. On the anterior side there are fewer muscles with short lever arms to the cervical spine. The muscle closest to the spine and the only one to attach on all cervical levels (except C1) is the Longus Colli. Postural functions of the Longus Colli and dorsal cervical muscles are complementary. They form a sleeve which encloses and stabilizes the cervical spine (31).

Altered patterns of muscle recruitment in both the cervical spine and shoulder girdle are features of chronic WAD, as measured with electromyography (EMG) (15, 25, 29, 30, 36-39). Greater perceived disability among patients with neck pain accounts for greater activation of the superficial, instead of deep cervical muscles (37). Deficits in the motor system have not only been reported in people with considerable disability, but also in those with persistent mild symptoms after a whiplash injury (15). However higher levels of pain are
associated with greater delays in the activation of the deep cervical flexors during rapid flexion of the shoulder (38). Contrary to asymptomatic individuals, multidirectional isometric contraction in women with WAD is reduced and less defined in patients with WAD confirming a disturbance in the neural control of the deep semispinalis cervicis muscle (39). Similarly, reduced head steadiness during isometric low load neck flexion in people with chronic WAD, especially related to severe pain and dizziness, compared to those with non-traumatic neck pain has also been reported (29). A tendency of higher muscle reactivity in the Trapezius in response to the exercise in people with WAD grade 2 compared to those with insidious onset has also been shown (36). Furthermore, people with neck pain of both whiplash and insidious origin have higher activity in the sternocleidomastoid during craniocervical flexion compared to control subjects without neck pain (25).

Peterson et al, using real time ultrasound with post-process speckle-tracking analyses, suggest altered interplay between deep and superficial cervical flexors, and between deep and superficial extensors during arm elevations in people with chronic WAD compared to controls without neck pain (24, 40). Interactions involved elongation of the deepest muscles in many of the WAD participants indicating that stabilization of the cervical spine did not occur in these individuals (41). Landén Ludvigsson et al (not part of this thesis) analysed different depths of the upper Trapezius muscle during an unloaded small scapular elevation task (shouldershrugging) with the same technique, which also indicated different muscle behaviours between people with chronic WAD and controls without neck pain. Controls used the superficial section of the Trapezius more than the deep section, whereas people with WAD used all three depths equally (23).

In people with WAD, muscle fatty infiltrates on magnetic resonance imaging (MRI) of the deep extensors (Multifidus and Semispinalis Cervicis) have been observed (42). Fatty infiltrates were however not features of insidious-onset neck pain as tested in women with persistent non-traumatic neck pain (43). Fatty infiltrates are significantly higher after 3 and 6 months in people with moderate/severe WAD compared to those recovered or with mild problems (44).

In both WAD and neck pain without traumatic onset also muscle fibers change over time. Transformation from slow to fast twitch muscle fibers (i.e. type I to type II fibers) has been observed for both deep and superficial cervical flexor and extensor muscles (45). Type I
fibers are characterized by low force/power/speed production and high endurance, whereas type II are the opposite.

**Chronic WAD**

An international review reports that 50% of those who suffer a whiplash trauma will report ongoing neck pain one year after their injury (4). There is often little spontaneous improvement beyond 3 months post-whiplash trauma (9, 46), and as mentioned there is no evidence of effective treatment (6, 7), which could be reasons why the definition chronic is the term still used to define this population with persistent symptoms in research. An alternative definition like for instance “longstanding” is not even an optional Medical Subject Head term (MeSH) on the large medical data base Pubmed (accessed 160225).

Both physical and psychosocial factors have been associated with the transition from acute to chronic WAD (15, 16, 19). In a recently presented prediction rule, high initial neck pain appears to be the only factor which consistently predicts poor functional recovery after a whiplash trauma across studies (47), but several factors have been proposed. Another recent review of risk factors for persistent problems following acute whiplash injuries also points out pain in connection to the trauma, especially high (>5.5 out of 10) baseline pain intensity, but also pre-injury neck pain. Other factors suggested in this review were having WAD grade 2 or 3 according to QTF, scoring high disability, (>14.5 out of 50) on the Neck Disability Index (NDI) (48) and being female (49). Postinjury psychological factors such as passive coping strategies, depressed mood, fear of movement and catastrophizing may however also be predictors of persistent disability (4), but the role of catastrophizing is conflicting (50-52). There also seems to be a role of vehicle crash-related factors. A higher risk of persistent disability if the head was rotated or inclined at the time of the impact (53), as well as direction of impact where rear-end impacts seem to cause more symptoms (54) have been reported. The impact can also be modified by impact awareness (55). However other factors like speed and weight of the vehicle are suggested not to play a major role (56).

**Treatment of WAD**

The most recent Cochrane report on conservative treatments for WAD grade 1-2 concludes that the overall methodological quality of studies is poor, and there is no clear evidence of any
effective treatments for either acute, subacute or chronic WAD (6). Only one small study (n=33) on chronic WAD was included in this report showing no difference between two active approaches (57). The most recent Bone and Joint Decade Task Force (2000-2010) on Neck Pain and its Disorders also states that it’s not possible to conclude what the most effective non-invasive treatment for chronic WAD is (7). Neither is there any clear evidence supporting behavioural therapy for persistent non-specific neck pain with or without radiating pain (58). However since persistence of symptoms in individuals with WAD has been attributed to both psychosocial and physical factors (16) it is not unreasonable to assume that a behavioural approach may be beneficial. Invasive treatment, such as radiofrequency neurotomy is reportedly effective for patients who have pain arising from the facet joints, however nerves recover and the procedure needs to be repeated and the long term effect is unknown (59, 60). Furthermore, both the test procedure, i.e. nerve blocks, and radiofrequency neurotomy are technically complex and the procedure is only effective in some patients (60). Behavioural interventions combined with exercises/physical activities are recommended for patients with long-standing neck pain by the Swedish Council on Health Technology Assessment (SBU) and the treatment can be performed by a physiotherapist in primary care since treatment outcome regarding pain is not better if treated by a multimodal team (61).

Exercise and physical activity

Physical activity can broadly be defined as any bodily movement generated by skeletal muscles resulting in energy expenditure. The terms physical activity and exercise are used interchangeably and are often viewed in nonspecific terms that include activities varying in type, intensity, and frequency. Exercise is a physical, biochemical, and social activity that can be manifested in a variety of forms, with the purpose of training or developing the body to promote physical health, specifically: flexibility, endurance, coordination, and relaxation (9). General physical activity and to stay active are common recommendations in the treatment of individuals with WAD in Sweden, although it has not been evaluated in chronic WAD (3).

Neck-specific exercise

Based on the observed changes in neck muscle behaviour in WAD, neck-specific exercises focusing on the deeper cervical muscles may be a feasible way of improving functioning. It
has previously been reported to be beneficial in chronic non-specific neck pain (62-65), cervicogenic headache (62, 66) and upper limp pain (62). Most of these studies have focused mainly on the deep anterior muscles however. A pilot study of 10 women with chronic WAD reports that the fatty infiltrates in the cervical Multifidus muscle can be reduced after 10 weeks of neck exercise (67). Schomacher et al. report that resistance applied to the head in extension activates both superficial (Splenius Capitis) and deeper neck muscles (Semispinalis Cervicis) (68). Äng et al. also conclude that early neck-specific and shoulder exercise interventions can be beneficial in reducing occurrence of neck pain in air force helicopter pilots (69). The most recent Cochrane report about exercise for mechanical neck disorders concludes that there are only temporary and benign side effects, and that exercise can be considered a safe intervention. There most likely is a role for exercises in the treatment of chronic neck pain and cervicogenic headache if focused on the neck and scapula region. However there seems to be no benefit for upper extremity stretching and strengthening exercises or a general exercise program (70).

Two studies have compared motor control training of the neck muscles with endurance and strength training of the neck muscles in chronic non-specific neck pain (65) and subacute WAD grade 1 and 2 (71). Both studies reported a reduced average intensity of neck pain and NDI score in both groups with no between-group differences, however there were no control groups. Randomized controlled trials (RCTs) of neck-specific exercise in chronic WAD are sparse. In the subacute stage (< 3 months’ duration) individually tailored supervised training, aimed at increasing cervical range of motion, cervical muscle endurance, stabilization, coordination and overall functional capacity has been suggested to be more favourable than home exercise, with a more rapid improvement in self-efficacy and pain disability (72). In one study of WAD grade 2 patients with more than 3 months’ duration, exercises of both the deep anterior and dorsal cervical muscles were part of a treatment protocol also including mobilization. The treatment protocol was better regarding pain reduction compared to a self-management program (booklet) (73). Another RCT included patients with WAD grade 1-3, of more than 3 months’ duration, where all participants received 3 advice sessions. In addition the experimental group participated in 12 exercise sessions over 6 weeks. The exercise was aimed at both endurance, coordination and fitness and a cognitive behavioural approach was used, but whether the exercise included neck-specific exercise is unclear. The exercise group had better outcomes in pain bothersomeness both at the 6 week and 12 month follow up. However pain and disability was only better at the 6 week, but not at the 12 month
follow up (74). High levels of baseline pain intensity were associated with greater treatment effects at 6 weeks and high levels of baseline disability were associated with greater treatment effects at 12 months (74). In chronic WAD grade 1-2 (> 3 months’ duration) simple advice is reportedly equally effective as a more intense and comprehensive physiotherapy exercise program with cognitive-behavioural therapy strategies (75). A pilot study from the study sample in this thesis, comparing neck-specific exercise with being on a waiting list revealed improvements in both neck disability, self-efficacy and health for the neck-specific group (76). No other RCTs have evaluated the effect of neck-specific exercise without the addition of other treatment approaches in chronic WAD grade 2 and 3, with > 6 months’ duration.

**Cost-effectiveness**

With an aging population, increasing demands on public health services and limited resources it is important to consider the cost-effectiveness of treatments given. Since the incidence of whiplash injuries has increased in the western world (2) costs for the treatment of WAD will most likely also increase. Costs associated with WAD are mostly attributed to health service costs for people with chronic symptoms and to the subsequent loss of work (2, 77). Yet a recent review of the literature failed to find any cost-effectiveness evaluations of treatment in chronic WAD (78).

Generic health-related quality of life (HRQoL) measurements are used in cost-utility analyses, but there is no absolute consensus on which measurement to use, even though the Euroqol 5 Dimensions (EQ 5D) (79) and Short-form 6 Dimension (SF 6D) (80) are commonly used. The results may however depend on the measurement used, since the SF-6D is better at detecting small changes in health and is more sensitive to changes in higher scores, whereas the EQ-5D is more sensitive to changes in lower scores (81). Furthermore, lately also disease-specific measurements like the NDI have been suggested as relevant options (82), since disease specific measurements are generally more responsive (83). The NDI has however not been used in cost-effectiveness analyses in WAD.
Outcomes

Functioning and Disability

In this thesis, functioning and disability were defined based on concepts defined by the World Health Organization (WHO) in the International Classification of Functioning, Disability and Health (ICF) (84). Functioning is an umbrella term which covers all body functions, activities and participation. Body function equals the physiological functions of the whole body system (including pure psychological functions, i.e. brain functions and the mind). This also includes for instance pain and sleep. Activity is the performance of a certain task by an individual, for instance getting dressed and participation is involvement in a life situation, for instance working.

Disability is the opposite (negative) term to functioning, and serves as an umbrella term for impairments, activity limitations or participation restrictions. Impairments is the term used to describe problems (deviation or loss) in body function such as loss of muscle function and structures (the neck). Activity limitation is the difficulty the individual might have performing a certain task. Participation restrictions are problems individuals might experience in involvement in life situations (i.e. based on the personal experience of the individual).

Pain

A commonly used definition of pain is the definition by the International Association for the Study of Pain (IASP) (85): “An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage”. Neck pain can be defined as “pain located in the anatomical region of the neck, with or without radiation to the head, trunk and upper limb” (86).

In the acute phase most neck pain is most likely nociceptive. After a whiplash trauma without recovery, the nociceptors can become hypersensitive and react excessively on otherwise non-painful mechanical stimuli. They can keep on firing nociceptive impulses towards the central nervous system either by outlasting the initiating input or by requiring a low-level peripheral drive to maintain the input (87). The dorsal horn neurons may then become hypersensitive, also progressing on to altered sensory processing in the brain and malfunctioning of descending pain-inhibitory mechanisms (87, 88) (central sensitization, or
The presence of hypersensitivity can influence outcomes in physical rehabilitation in chronic WAD where those having both widespread mechanical and cold hyperalgesia, which can be signs of central sensitization, showed least improvement in one study (73). Van Oosterwick reported impaired endogenous pain inhibition during submaximal exercise in people with chronic WAD (89). Due to the increased responsiveness of the central nervous system in some individuals with chronic WAD, recommendations are therefore to be cautious with the additional or accelerated source of nociception, since it may sustain or enhance the process of central sensitization (88). This however not quite in accordance with common behavioural approaches like the graded treatment approach, where focus is on success in for instance an exercise progression, despite possible pain provocation (90).

**Psychosocial factors, focusing on self-efficacy**

Chronic WAD involves a variety of symptoms with considerable overlap between organic and psychosocial origins. Baltov et al. found, based on interviews, that psychosocial factors played a role in distress and return to work, but not in pain and disability in chronic WAD (91). However Bunketorp suggested that the most important predictor of persistent disability in people with subacute WAD is low self-efficacy (92), and later showed that self-efficacy was the most important predictor of persistent WAD one year after the injury (93). Denison et al. found that self-efficacy is a better predictor of disability than both fear avoidance and pain intensity in a primary health care sample of patients with subacute, chronic, or recurring musculoskeletal pain. This result was shown in two different samples (94). Söderlund and Åsenlöf compared a group of people with acute whiplash trauma with a group of people with other acute traumas, for instance extremity sprains or fractures. They found that fear of movement proved to be a mediator between pain intensity and pain-related disability in the musculoskeletal-injury group, but in the WAD group self-efficacy was a mediator (95). There is a connection between disability and self-efficacy. Disability may be perceived even without tissue injury and with low pain intensity in an individual who lacks self-efficacy (92, 96).

Self-efficacy is a concept used in different behavioural models, but is originally mostly associated with Social Cognitive Theory. Self-efficacy is described as “*the conviction that one can successfully execute the behaviour required to produce the outcomes*”, by the psychologist Albert Bandura, who developed the concept (97). Self-efficacy beliefs are thus based on an individual’s subjective perception of reality, and not on the objective facts. There
are four major ways of improving self-efficacy: The one with the strongest influence on self-efficacy is through “mastery experience”. That means enabling an individual to succeed in increasingly challenging performances of a behaviour that is attainable and desired. Another way is through “Social modelling” which means that the individual is shown that other people like themselves can perform a certain task. The third way is “Improving physical and emotional states”. This can include efforts to reduce stress and depression while building positive emotions. Finally “Verbal persuasion” means telling the individual that he or she can manage to do a certain thing. Strong encouragements can improve confidence enough to prompt the first efforts toward changing a behaviour (97), for instance starting up an exercise program. To improve self-efficacy it is thus important to put up goals that are achievable, rather too simple than too difficult, and small partial goals may be needed. It is also important to try to focus on success primarily, rather than focusing on failure. Regardless of the initial level of self-efficacy before entering an exercise program, self-efficacy has been shown to improve over the scheme for completers, whereas it tended to deteriorate for drop-outs (98).
RATIONALE OF THE THESIS

As previously stated, 50% of those who suffer a whiplash trauma will report ongoing neck pain one year after their injury (4), and yet there is a lack of evidence for any effective treatments for WAD (6, 7). Since altered neck muscle behavior is a feature of chronic WAD (23-25, 30, 36, 37), neck-specific exercise may be a feasible intervention that has reportedly had good effect on other neck pain conditions (62, 99). But there is also a psychosocial aspect and exercises/physical activities combined with behavioural interventions are recommended for patients with longstanding neck pain in Sweden (61), even though there is a lack of evidence for this approach in chronic WAD.

Even though the majority of people with chronic WAD also report contemporary symptoms like headache and/or upper extremity symptoms (43, 44) the few available studies of exercise and WAD most often seem to exclude people with neurological signs (grade 3 according to the QTF). No previous RCTs including participants with WAD grade 3, have evaluated the effect of neck-specific exercise without a combination of other treatment approaches, nor evaluated prescription of general physical exercise in chronic WAD, nor compared prescription of physical activity to therapist-led neck-specific exercise with or without a behavioural approach, which was examined in this thesis.
AIMS

General aim and hypothesis

The general aim of this thesis was to evaluate the effect on self-reported pain, disability/functioning, and self-efficacy of three different exercise interventions in chronic WAD grade 2 (without neurological symptoms) and 3 (with neurological symptoms) and to determine the cost-effectiveness of these interventions. The hypothesis was that neck-specific exercise would have a better effect on pain and disability than prescription of physical activity, and that the addition of the behavioural approach would result in superior improvements in disability and self-efficacy than neck-specific exercise alone. Another hypothesis was that neck-specific exercise, with or without a behavioural approach, would be cost-effective and may contribute to increased general health in chronic WAD, grade 2 and 3.

Specific aims

The specific aims were:

- To evaluate the effect on self-rated pain, functioning/disability and self-efficacy of three interventions in chronic WAD management, grade 2 or 3: physiotherapist-led neck-specific exercise, physiotherapist-led neck-specific exercise with a behavioural approach, or prescription of physical activity.

- To explore determinants of clinically important disability and/or pain reduction in people with chronic WAD, grades 2-3, following exercise interventions.

- To analyse cost-effectiveness following three exercise interventions in chronic WAD grade 2 and 3: physiotherapist-led neck-specific exercise alone or in combination with a behavioural approach and the prescription of physical activity comparing two different health related quality of life measurements: the EQ-5D, the SF-6D and the disease-specific NDI.
MATERIAL AND METHODS

Design

All papers in this thesis are based on a multicentre, prospective, randomized controlled clinical trial, with assessor and group allocation blinding (Clinical Trials.gov, no NCT01528579) (100, 101). It is an effectiveness trial measuring the degree of beneficial effect under clinical settings. Outcomes used in this thesis are specified further on (please see “Outcomes” and table 4), whereas other outcomes specified in the protocol are analysed elsewhere. An overview of the analyses of the papers in this thesis can be found in table 1. Due to the nature of the interventions participants and treating physiotherapists were unable to be blinded to the interventions.

Table 1. Overview of analyses of the included papers in the thesis

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short term follow-up of self-reported disability, pain and self-efficacy (change scores up to 6 months) Between-group comparison without imputations</td>
<td>Factors associated with clinically relevant improvement in disability and pain after 3 and 12 months</td>
<td>Long term follow-up of self-reported disability, pain and self-efficacy (change scores up to 24 months) Between-group comparison with imputations</td>
<td>Cost-effectiveness, between-group comparison during 12 months from a societal and health care perspective</td>
</tr>
</tbody>
</table>

Participants, recruitment and randomization

Participants were recruited between February 2011 and May 2012 and all participants received verbal and written information about the study. In the first step of recruitment to the study, a large number of letters (n=7950) were sent to potential participants aged 18-63, enquiring about their interest to participate. They were identified from health care records in six counties from mainly primary care, but also emergency, orthopedic and neurosurgical departments, having sought care in the preceding 6-36 months due to whiplash associated diagnoses. In this thesis chronic WAD is defined as having at least 6-month duration of symptoms, which is a definition also previously used in WAD research (88, 102). It was chosen to ensure minimal change could be expected without any intervention. Since in chronic states, the clinical experience of the project leaders, was that the whiplash diagnosis is often changed to that of just cervical pain, with or without radiculopathy, letters to people with such diagnoses were also sent. The letters included basic inclusion/exclusion criteria.
(i.e. a whiplash injury in the preceding 6-36 months, reported to be the onset of current symptoms, no previous neck trauma with unresolved symptoms, no previous neck surgery or ongoing malignant disease, no severe psychiatric disorders, drug abuse, or difficulties in understanding the Swedish language). All inclusion/exclusion criteria are listed in table 2. They were also asked to fill out the NDI and Visual Analogue Scale (VAS) (103) of average neck pain in the preceding week. Two reminders were sent to those with a whiplash diagnosis, and one to those without. The majority were not eligible (n=7531). A telephone interview was undertaken by one of the project leaders with the remaining 419 potential participants to confirm inclusion/exclusion criteria, and evaluate if further review of medical files was required to determine eligibility and for further information to the potential participant. In cases where eligibility was uncertain, medical files were checked with consent from the potential participant. Eligible participants then attended a physical examination to confirm WAD-grading (2 or 3) where informed consent was also obtained. WAD grade 2 was defined as previously described. In this thesis the criteria of neurological signs (WAD grade 3) was met if two or more of the neurological tests in the physical examination rendered positive observations in the same dermatome/myotome C4-C8: sensibility (using both brush and pinwheel), strength, reflexes, and provocation or relief of current arm pain by neck traction in lying or Spurling’s test of compression in sitting. This classification has previously been used and found reliable in classification of neurogenic pain and dysfunction in the neck/shoulder region together with a modified Pain Drawing (104). In our study arm symptoms were also to be reported in either a Pain Drawing, or in reply to the question “Have you got arm pain/ numbness/prickling sensations in connection with your neck pain” at least “from time to time” (which equals grade 2 or more often in a 5 graded scale from 1 “never” to 5 “all the time”) without other known causes of the arm symptoms, to fulfill the criteria.

A further 203 individuals were excluded and finally 216 participants were included (figure 1) including 142 (65%) women and 74 (35%) men with a mean age of 40.5 years (range, 18 to 63 y, SD 11.4). Baseline outcome measurements were collected before allocation. Allocation from a randomization list was made by an independent researcher, not otherwise involved in the study, who also put the individual results in sealed envelopes for further distribution directly to the treating physiotherapists.
Table 2. Inclusion and exclusion criteria

Inclusion criteria
- Age 18-63 years
- Whiplash Associated Disorder (WAD) grade 2-3
- Whiplash-type neck trauma at least 6 but no more than 36 months ago with persistent neck pain (>20mm on 100mm Visual Analogue Scale (VAS) and/or disability >10/50 on the Neck Disability Index (NDI))

Exclusion criteria
- Myelopathy, spinal infection or tumour
- Generalized or more dominant pain elsewhere in the body.
- Direct head trauma and/or unconscious/loss of memory in connection to the trauma
- Earlier neck trauma with persistent neck problems
- Neck pain causing more than one month’s work absence the year before the trauma
- Previous surgery in the cervical spine
- Ongoing malignant disease
- Diseases or other injuries that might prevent full participation in any of the interventions.
- Diagnosed severe psychiatric disorder
- Known drug abuse
- Insufficient knowledge of the Swedish language.

Of the total 216 participants included in this study, 57% (n=122) were classified as WAD grade 2 and 43% (n = 94) as grade 3. Participants with WAD grade 3 reported higher baseline disability (NDI, grade 2; 15.6 (SD 6), grade 3; 18.2 (SD 7.1), p<0.01) and neck pain (VAS, grade 2; 39 (SD 24), grade 3; 47(SD 24), p=0.02). The also reported lower pain-related self-efficacy (Self-efficacy Scale, grade 2; 157 (SD 34), grade 3; 140 (SD 37), p=0.001).

Apart from a slight, but significant difference in age and gender, there were no differences between the three intervention groups at baseline (table 3). There was however no correlation between age (all Rs <0.11, P>0.19), gender (all Rs <0.12, P>0.09), and any of the outcomes. The drop-out rates were: at 3 months 12%, at 6 months 20%, at 12 months 21%, and at 24 months 43% (figure 1). There was no difference in the baseline outcomes (p > 0.27 for all) between participants who dropped out and those who completed up to 12 months, except drop-outs at 12 months were somewhat younger (age 37 (SD11) vs 41 (SD11), p=0.04). At 24 months drop-outs had reported more baseline disability than those who completed the study (NDI 17.9 (SD 6.9) vs NDI 15.7 (SD 6.5), p < 0.04), but their improvements up to 12 months were the same (p =0.94).
Figure 1. Flow chart of participants

*Whiplash injury in the preceding 6-36 months, reported to be the onset of current symptoms, unconsciousness/loss of memory in connection to the whiplash injury, previous neck trauma with unresolved symptoms, previous neck surgery, ongoing malignant disease, severe psychiatric disorders, drug abuse, difficulties understanding the Swedish language.
Interventions

The interventions were: 1. physiotherapist-led Neck-Specific Exercise (NSE), 2. NSE with the addition of a behavioural approach (NSEB) or 3. Prescription of Physical Activity (PPA). The participating physiotherapists, who worked in primary care, were used to treating neck pain patients. As much as possible the physiotherapists were selected and matched to work within their field of knowledge and interest and those in charge of the NSEB group also generally had a special interest/experience and/or further education within behavioural treatments. The interventions were offered as close as possible to the participants’ home or work place. Since the study was conducted in 6 counties, many physiotherapists (n=69) served as treating physiotherapist. In a few cases, in small towns where not enough physiotherapists were available, the treating physiotherapists delivered more than one intervention (n=3 physiotherapists). The physiotherapists conducting the interventions were provided with standardized oral and written information about their interventions, and those delivering the physiotherapist-led exercise interventions received one day of standardized theoretical and practical training from the project leaders. Throughout the study all physiotherapists were encouraged to contact the project leaders if in need of any further guidance. All interventions started with an individual examination of the participants by their treating physiotherapist. All three interventions were undertaken over a 12 week period. Participants were urged to refrain from having any other physical treatments for their neck disorder at least during the first 6 months of participating in this study. Time frames and components of the interventions are specified in figure 3.

**Physiotherapist-guided neck specific exercise (NSE)**

In this group, participants undertook supervised exercise, and received basic information about the musculoskeletal anatomy of the neck, relevant to the exercise. The importance of good posture was also emphasized to further facilitate deep cervical muscle function (105). The exercise program consisted of two physiotherapy sessions weekly where the initial focus was to guide activation of deep cervical muscles without activation of superficial ones through gentle unresisted isometric cervical flexion, extension and rotation exercises in supine (figure 2). The participants then practised daily at home with a recommended starting point of about 3 sets of 5 repetitions of each exercise. Exercise was then progressed in each direction with low resistance, increasing the exercise parameters towards 3 sets of 10 repetitions in supine and thereafter in sitting positions. After 2-3 weeks of guided home
exercise, resisted exercise was gradually introduced in the gym. Resistance and number of repetitions using a weighted pulley for head-resistance (figure 2), or guild board was used with a focus on endurance training progressing to higher repetitions (if possible up to 3 sets of 30 repetitions) without further pain provocation. Although a standardised framework of exercises was followed, progression was tailored to each individual according to their symptomatic response and capability. The exercise program could also include one exercise each for the lower back, abdomen, and scapulae, as well as stretching exercises, if considered appropriate for that individual. Towards the end of the 12 week exercise period participants were encouraged to continue exercise at home by providing them with resistive exercise bands and a written individualized exercise program also including prescription of general physical activity. Time frames of the exercise program are described in figure 3, and the exercises are further described at Academic Archives on line (106).

Physiotherapist-guided neck-specific exercise with a behavioural approach (NSEB)

The exercises in the NSEB group was identical to that undertaken by the NSE group, including the recommendations of continuous exercise post-intervention. However it was initially progressed slightly slower to accommodate the additional behavioural component (figure 3). Participants were encouraged not to focus on temporary neck pain increase, in accordance with the concept of behavioural graded activity, but provocation of radiating pain/neurological symptoms was avoided. Apart from not focusing on pain, participants were also encouraged to
MATERIAL AND METHODS

take responsibility for the exercise progression, with the physiotherapist as a coach, in an operant-conditioning behavioural approach which is also part of a behavioural graded treatment approach (90). The focus was thus on success in exercise progression, not pain reduction. The operant-conditioning model states that reinforcement of pain behaviours leads to their maintenance (107). The participants also received introductions to basic behavioural intervention, led by the physiotherapist, underpinned by concepts mainly, but not solely, from Social Cognitive Theory (108) and Transtheoretical Theory (109). The behavioural approach was designed based on the experience of physiotherapists/researchers with further education and experience within the field of behavioural treatments, to be basic and manageable by physiotherapists in primary care.

One of the most well-known concepts in Social Cognitive Therapy is self-efficacy, and one aim with this intervention was to improve self-efficacy considering the four major ways previously described: to achieve “mastery experience” focus was on success in an increasingly challenging exercise program, and to improve “physical and emotional states” awareness of the influence of thoughts on behaviour, relaxation and breathing exercises were introduced. “Verbal persuasion” was used by the physiotherapist by encouraging the participant to reach a goal or do the exercises. The fourth way, “social modelling”, may have been more difficult to achieve since there may not have been other patients with WAD who had come further in their rehabilitation available at all physiotherapy gyms.

Self-efficacy is also a concept used in the Transtheoretical Theory, but the core concept is a process of states of change. First of all participants need to be prepared to make some changes to change a behaviour, e.g. starting to exercise, which can be expected when enrolling in an exercise study. Participants participated in pain management education, including oral education regarding physiological and psychological aspects of pain, with an emphasis on how pain and disability can be sustained even when the injury itself has healed. This was aimed at steps of the process of change, which includes increased “awareness” about causes of chronic pain, which may lead to “relief”, and “self-revaluation”. A shorter version of this education was repeated later on during treatment, in order to be able to correct misinterpretations of the participant, and to reinforce the understanding. Performing the exercise also helps the self-revaluation of the participant as a more active person. Personal goals were also set, to encourage the participant that the participant can change and achieve goals, as a way of self-revaluation and also the belief that he/she can change, and may
MATERIAL AND METHODS

substitute less healthy behaviours, such as avoidance of being active, for more healthy behaviours such as exercising. The time frames of the behavioural components are described in figure 3.

**Prescription of Physical Activity (PPA)**

Participants in this group first had a short motivational interview conducted by the physiotherapist. In a motivational interview the physiotherapist listens, tries to understand the participant’s perspective and emphasizes that the participant finds his/her own answers and decisions regarding change, i.e. exercise, which is what the participant has enrolled to do in the study. It can therefore be expected that the participants in this study already, by actively showing their interest in the study and by enrolment, were either ready to act or are at least not disinterested. Motivational interviewing is not used to try to get a participant to do something that he/she does not want to do, but rather stimulating the participant to change his/her own conditions in a respectful manner. Based on the discussions within the interview and the physical examination the participants were prescribed individualized physical activity (110). The aim of this prescription was to increase overall physical activity, either with activities performed outside the healthcare system, for instance at public gyms, or with individualized home exercise. Common examples of exercise could be gym classes, Nordic Walking, cross-trainer exercise or walks in combination with individual home exercises. Neck-specific exercises were not prescribed in this group. One follow-up visit or phone call was encouraged to make sure the participant felt the activities were suitable and possible to perform (figure 3). Consistent with the approach taken with the NSE and NSEB groups, participants were encouraged to continue exercising after the 3 month intervention was over.
**Neck-specific exercise (NSE)**

**Week 1.** Individual examination, facilitation of deep cervical muscle activity through unresisted neck-specific exercise in supine, provision of basic information regarding motor function of the neck, and information about the purpose of therapeutic exercise

**Week 2-3.** Exercise progressed to isometric neck-specific home exercise first in supine, then in sitting, and thereafter introduction to gym exercises and instruction in postural control

**Week 3-12.** Ongoing progressive neck-specific gym exercising, home exercises with elastic resistive bands.

**Week 12.** Recommendation of continuous neck-specific exercise following completion of intervention period and prescription of physical activity

---

**Neck-specific exercise with an additional behavioural approach (NSEB)**

**Week 1.** Individual examination, facilitation of deep cervical muscle activity through unresisted neck-specific exercise in supine, pain management education, activity goal setting, relaxation exercises

**Week 2.** Instruction of postural exercise, continued unresisted neck-specific exercise.

**Week 3.** Neck-specific home exercise progressed to isometric neck-specific home exercise first in supine then in sitting, awareness of the influence of thoughts on behaviour

**Week 4.** Commencement of graded neck-specific exercise in gym, activity based goal specific exercise

**Week 5-12.** Ongoing graded neck-specific gym exercising, neck-specific home exercise with elastic resistive bands

**Week 5.** Discussion of personal activity level/pacing, breathing exercises

**Week 7.** Reinforcement of pain management education

**Week 8.** Follow-up of activity goal setting

**Week 10.** Reinforcement of strategies to handle relapse/periods of worsening

**Week 11-12.** Recommendation of continuous neck-specific exercise following completion of intervention period and prescription of physical activity, follow-up activity goal setting

---

**Prescription of physical activity (PPA)**

**Week 1.** Individual examination, motivational interview, prescription of individualized physical activity

**Week 2-12.** Physical activity outside health care system, possibility of 1 follow-up with physiotherapist

---

**Figure 3. Specific components of interventions and timeframes.**

**Data collection**

Data was collected at baseline and after 3, 6, 12 and 24 months. All outcome measurements were collected and registered by research staff blinded to intervention group allocation. At all time points, except the 24-month follow-up, both questionnaire data and results of physical examinations, performed by blinded test leaders, were collected. The results of the physical
examinations will be reported elsewhere. At 24 months, only questionnaires were collected. Participants completed questionnaires at home, including background variables such as generic data, activity level, measured with International Physical Activity Questionnaire (IPAQ) (111), and other neck related questions. The questionnaires also included specific questionnaires used as outcomes in the different papers (table 4). One questionnaire, the Patient Specific Functional Scale (PSFS) (112) was completed verbally with the test leader, blinded to the interventions, who undertook the initial screening physical examination, and the physical examination follow-ups. Exercise data was also collected from the treating physiotherapists up to 3 months post-inclusion, and exercise diaries from the participants were collected in sealed opaque envelopes up to 6 months post-inclusion. Thereafter participants were asked in the questionnaires to estimate their adherence to their post-intervention prescribed exercise on a 4-point scale: full, fair, some, or no adherence.

Outcomes

Outcomes used in each paper are specified in table 4.

Main Outcome

The Neck Disability Index (NDI)

The NDI was originally developed by Vernon and Mior (48) from the valid and reliable Oswestry Low Back Pain Disability Questionnaire (113). The NDI is the most widely used questionnaire for disability due to neck pain (114, 115). It is highly reliable and strongly internally consistent, as documented in several studies, and has strong and well-documented validity (114), and been found valid and reliable in chronic neck pain (116) and in subacute WAD (117). It has been translated into Swedish and found reliable (116). It has five items that are classified to impairments, three to activity limitations and two that are indicative of participation restrictions (115). The 10 items are; pain intensity, personal care, lifting, sleeping, driving, recreation, headache, concentration, reading and work. Each items has 6 possible answers which are scored 0 (no limitations) to 5 (major limitations) and are summed up to yield a total score of 50. The index can be used either as a score of maximum 50, which is most commonly used, or as a percentage out of 100% (114). There is some disagreement whether it has a one factor (118) or two factor structure (115, 117), and whether these two
factors correlate or not. In WAD a two-factor structure is suggested, where one of the factors is labelled “pain and interference with cognitive function” (neck pain, headache, sleeping, concentration and reading), and the other one is labelled “functional disability” (work, personal care, driving, lifting, recreation). The two factors have been found to correlate (117).

**Secondary Outcomes**

*The Patient Specific Functional Scale (PSFS)*

In the Patient-Specific Scale (PSFS) each participant lists individual specific activities they are unable to do or have difficulty performing as a result of their condition. There are usually three different activities. The activities are ranked on scale from 0 (unable to do) to 10 (functional level equal to preinjury status). The final score is determined by averaging the three activity scores. Higher scores represent a greater level of function. Test-retest reliability has been demonstrated to be excellent (ICC 0.82) in a sample of patients with cervical radiculopathy (119), but was not reproduced in a larger sample (120). High ICC (0.92) has also been demonstrated for test–retest reliability in patients with neck dysfunction (112). The PSFS score tested for cervical radiculopathy is the one that has been found the most reliable and valid, since not enough information is available on other groups (121). In chronic WAD the PSFS has been reported to be the most responsive measure of disability (122). A self-reported version has been found well related to an interviewer administrated version, and thus found to suffice in criterion validity (121).

*The Visual Analogue scale (VAS)*

The complexity of pain makes it difficult to measure, but a common measurement used, both clinically and in research, is the VAS scale (123). It is a straight line, anchored by the extreme boundaries of the response to be measured (103). A horizontal line is recommended (124). Ogon et al found that data was normally distributed when the VAS was used horizontally, but not when it was used vertically (125). It is usually a line of 100 mm thus giving a range of 0-100. In this thesis pain and pain bothersomeness were the responses measured with a VAS-scale. Pain bothersomeness is suggested to be more responsive than pain intensity in chronic WAD (122). The anchor ends for the VAS scales were no pain = 0 and worst imaginable pain =100 (VAS-P), and = 0 not bothersome at all, to extremely
bothersome = 100 (VAS-B). The VAS-scale has been found reliable regarding test-retest and
terrater reliability. It has been found to be more sensitive to change than numerical rating
scales and there are some conflicting results regarding correlation between these scales
(126). According to the Initiative on Methods, Measurement and Pain Assessment in Clinical
Trials (IMMPACT) recommendations (127), substantial improvement of pain intensity, or
treatment success, is suggested when a \( \geq 50\% \) reduction of pain is achieved.

*The Self-Efficacy Scale (SES)*

Self-efficacy was evaluated using the Self Efficacy Scale (SES)(128). The SES has been
translated into Swedish and has previously been used in studies of WAD. The Swedish
version has been found reliable in patients with WAD both regarding internal consistency
(19, 50, 94, 129), and test-retest reliability (129). It was originally constructed on the basis of
Banduras concept of self-efficacy and it was originally developed for and tested on low back
pain patients (130). It contains 20 questions, based on activities that vary in performance
difficulty and have high item-total correlations. These questions relate to everyday activities,
mainly physical and psychosocial, but also to cognitive skills. Patients are asked to rate on a
10 point scale how confident they are in their ability to successfully complete each task
despite their pain: for instance going shopping, shovelling snow, driving the car, eating in a
restaurant, watching television, visiting friends, raking leaves, writing a letter, doing a load of
laundry, working on a house repair, concentrating on a project, washing the car, riding a
bicycle, or going for a walk. Each item is scored from 0 = not confident at all, to 10 = very
confident, and generates a total score from 0-200. Higher scores thus indicate higher self-
efficacy. In the Swedish version, translated by Denison et al (94), a few words in the
instruction (related to low back pain) were changed. These were instead “people who have
pain” to suit other groups than the initial low back pain group.
Table 3. Background variables and baseline outcomes scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>NSE (n=76)</th>
<th>NSEB (n=71)</th>
<th>PPA (n=69)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, female, n (%)</td>
<td>57 (75)</td>
<td>47 (66)</td>
<td>38 (55)</td>
<td>0.04</td>
</tr>
<tr>
<td>Age, mean (range) SD</td>
<td>38 (18-62)</td>
<td>40 (19-63)</td>
<td>43 (19-63)</td>
<td>0.03</td>
</tr>
<tr>
<td>Months since injury, mean (range)SD</td>
<td>19 (6-36)</td>
<td>20 (6-36)</td>
<td>20 (6-36)</td>
<td>0.69</td>
</tr>
<tr>
<td>Neck pain the year before accident, n (%)</td>
<td>12 (16)</td>
<td>6 (9)</td>
<td>12 (18)</td>
<td>0.25</td>
</tr>
<tr>
<td>Motor vehicle accident, n (%)</td>
<td>65 (86)</td>
<td>54 (76)</td>
<td>54 (82)</td>
<td>0.32</td>
</tr>
<tr>
<td>WAD grade 2/3, n (%)</td>
<td>49/27 (64/36)</td>
<td>33/38 (46/54)</td>
<td>41/28 (58/42)</td>
<td>0.08</td>
</tr>
<tr>
<td>Previous physiotherapy treatment, n (%)</td>
<td>58 (77)</td>
<td>52 (75)</td>
<td>45 (68)</td>
<td>0.44</td>
</tr>
<tr>
<td>Smoker, n (%)</td>
<td>17 (22)</td>
<td>8 (11)</td>
<td>12 (18)</td>
<td>0.22</td>
</tr>
<tr>
<td>Educational level, elementary, n (%)</td>
<td>4 (5)</td>
<td>6 (9)</td>
<td>6 (9)</td>
<td>0.44</td>
</tr>
<tr>
<td>Educational level, high school, n (%)</td>
<td>38 (50)</td>
<td>40 (57)</td>
<td>34 (51)</td>
<td></td>
</tr>
<tr>
<td>Educational level, university, n (%)</td>
<td>31 (41)</td>
<td>21 (30)</td>
<td>24 (36)</td>
<td></td>
</tr>
<tr>
<td>Educational level, other, n (%)</td>
<td>3 (4)</td>
<td>3 (4)</td>
<td>3 (4)</td>
<td></td>
</tr>
<tr>
<td>Use of analgesic drugs, n (%)</td>
<td>40 (53)</td>
<td>44 (62)</td>
<td>45 (67)</td>
<td>0.23</td>
</tr>
<tr>
<td>Employed n (%)</td>
<td>61 (80)</td>
<td>57 (80)</td>
<td>52 (75)</td>
<td>0.71</td>
</tr>
<tr>
<td>*General Health,*EQ-5D-score, median (IQR)</td>
<td>0.72 (0.69-0.76)</td>
<td>0.73 (0.23-0.80)</td>
<td>0.73 (0.66-0.80)</td>
<td>0.62</td>
</tr>
<tr>
<td>Physical activity level, IPAQ</td>
<td></td>
<td></td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>Physical activity level, low, n (%)</td>
<td>40 (56)</td>
<td>36 (56)</td>
<td>39 (58)</td>
<td></td>
</tr>
<tr>
<td>Physical activity level, medium, n (%)</td>
<td>12 (17)</td>
<td>8 (13)</td>
<td>15 (22)</td>
<td></td>
</tr>
<tr>
<td>Physical activity level, high, n (%)</td>
<td>19 (27)</td>
<td>20 (21)</td>
<td>13 (19)</td>
<td></td>
</tr>
<tr>
<td>Baseline outcome scores:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI, mean (SD)</td>
<td>16 (6)</td>
<td>17 (7)</td>
<td>17 (7)</td>
<td>0.47</td>
</tr>
<tr>
<td>PSFS, mean (SD)</td>
<td>4.4 (1.7)</td>
<td>4.5 (2.1)</td>
<td>4.6 (1.8)</td>
<td>0.80</td>
</tr>
<tr>
<td>VAS P, mean (SD)</td>
<td>40 (24)</td>
<td>45 (24)</td>
<td>42 (25)</td>
<td>0.51</td>
</tr>
<tr>
<td>VAS B, mean (SD)</td>
<td>49 (22)</td>
<td>50 (23)</td>
<td>48 (22)</td>
<td>0.87</td>
</tr>
<tr>
<td>SES, mean (SD)</td>
<td>150 (34)</td>
<td>153 (35)</td>
<td>147 (41)</td>
<td>0.52</td>
</tr>
</tbody>
</table>

NSE: neck-specific exercise group; NSEB: neck-specific exercise group with a behavioural approach; PPA: prescription of physical activity group; EQ-5D: Euroqol-5D health questionnaire; IPAQ: International Physical Activity Questionnaire; NDI: Neck Disability Index; P-VAS: Pain Visual Analogue Scale; B-VAS: Pain Bothersomeness Visual Analogue Scale; SES: Self Efficacy Scale; PSFS: Pain Specific Functional Scale; IQR: interquartile range; SD: standard deviation.

Clinical relevance of outcome measurements and adherence

The IMMPACT concludes that it is impossible to provide specific guidelines for determining whether a group difference is clinically meaningful or not, and that is has to be decided on a
MATERIAL AND METHODS

To evaluate not only statistical significance, but also the clinical relevance of the results, the proportion of participants reaching an important clinical improvement was thus evaluated. The Minimal Clinically Important Difference (MCID) is defined 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 the patient’s management” (132). The MCID of the NDI-score is suggested to be 3.5-5/50 points in non-specific neck pain (114), and was set at a cut-off score of at least 5/50 reduction defining clinically relevant improvement regarding disability in this study. To make sure it exceeded the measurement error (SEM) in this study population, Minimum Detectable Change (MDC) was calculated (119). The MDC is minimum amount of change that ensures the change isn't the result of measurement error. The calculation (1.96 x SEM x square root of 2) was done using the NDI scores from both the enquiry letters, and baseline questionnaire from 5 counties (n=180), rendering a MDC of 2.4/50 points. In the PSFS, the MCID is reported as 2.3 and the MDC is reported as 2.1 (119).

A reduction in pain intensity of ≥50% is suggested to indicate substantial improvements, or treatment success (127), and was used in this thesis to define clinically important improvement regarding pain. The MCID has not been established in the SES.

To make sure at least some of the intended interventions had been performed, in this part of the analyses only participants who had at least 50% attendance to the intervention sessions (data collected from the physiotherapists and from participant exercise diaries) were part of these analyses at 3-6 months (NSE 70%, NSEB 71% and 50% from the PPA group) (paper 1) (101). Participants who reported some to full adherence at 12 and 24 months were part of the long-term analyses (NSE and PPA 79%, and 69% from the NSEB group) (paper 3), (133).

Other measurements – potential factors associated with outcomes

In paper 2 (134) an analysis was made to examine whether the type of exercise intervention alone was a determinant of clinically important disability and pain reduction, or if baseline factors were more dominant factors associated with the outcome. The treatment variable can be included as a separate predictor in such a multivariable model (135) and was one of the factors included. Baseline factors considered as potential factors associated with the outcome’s improvement in pain and disability were selected based on clinical experience of
MATERIAL AND METHODS

two specialist physiotherapists with over 15 years’ experience managing patients following a whiplash injury (136) and on a review of the literature. The factors chosen were age, gender, patient expectations of symptom relief from each of the three interventions (Numeric Rating Scale 0-10, 0 = no expectations of help, 10 = best possible help, matched with each individual randomized intervention), dizziness (never/from time to time or more often), baseline neck pain intensity (VAS-scale 0-100), use of analgesics and WAD-grade.

To capture more complex concepts, scores from the following scores from the collected questionnaires (100) were also used:

Pain catastrophizing, which may affect the perception of and response to persistent pain, was measured using the Pain Catastrophizing Scale, (PCS) (137). This is a 13-item questionnaire that describes various feelings and thoughts that people experience when reflecting on their past pain experiences when in pain. Each item has a 5-point score from 0 = not at all to 4 = all the time, and it generates a total score from 0-52. Higher numbers indicate more pain catastrophizing. It is valid (137) and has been used previously in the evaluation of WAD treatment (138).

Kinesiophobia, or fear of movement and (re)injury, was measured using the Tampa Scale for Kinesiophobia, TSK-11 short form. The scale was designed to discriminate between non-excessive fear and phobia among patients with chronic musculoskeletal pain. It is an 11-item questionnaire where each item has a 1-point score from 1 = strongly disagree, to 4 = strongly agree, and it generates a total score from 11-44, with higher numbers indicating more fear of movement. It is considered valid and reliable, and has been used in the evaluation of WAD (139).

Other scores used, as previously described, were neck-related disability (NDI) (48), self-efficacy of performing various activities despite pain (SES) and level of physical activity (IPAQ) (111)

Also clinical tests were considered potential factors associated with the outcomes:

Cervical flexor and extensor muscle endurance were tested. Flexors were tested in supine, with the chin retracted. Extensors were tested in prone with a load of 2 kg for women or 4 kg for men applied to the head. The participants were instructed to lift their head just above the examination table and to maintain the test positions to the point of neck fatigue, radiating pain, or severe neck pain. For both tests, endurance was measured in seconds using a stopwatch (140).
Total active range of neck motion in sagittal and rotation in sitting was also measured in degrees, with a Cervical Range of Motion Device (CROM) (141). Active full range of motion in one plane was used since it is reported to be more reliable than a passive range of motion or motion to one side at a time (142).

Finally, the number of cervical levels of sensibility disturbance (using a brush or pinwheel for C4-C8 dermatomes) were measured. All data were collected at baseline prior to randomization (paper 2), (134).
Table 4. Measurements used in the different papers

<table>
<thead>
<tr>
<th></th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-reported:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>VAS current neck pain</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAS pain bothersomeness</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>PSFS</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Analgesics</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>SES</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>TSK-11 Short Form</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>EQ-5D</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>SF-6D</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>PCS</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>IPAQ</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Dizziness</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Patient expect symptom relief</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health care consumption</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td><strong>Generic:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical tests:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neck flexor endurance</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Neck extensor endurance</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Tot active sagittal range of neck motion (CROM)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tot active range of neck motion rotation (CROM)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAD grade 2 or 3</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Number of cervical levels of sensibility disturbance</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Register data:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production loss (Sick-leave)</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

NDI = Neck Disability Index, VAS = Visual Analogue Scale, PSFS = Patient Specific Functional Scale, SES = Self-Efficacy Scale, TSK-11 Short Form = Tampa Scale for Kinesiophobia, Short Form 11, EQ-5D = Euroqol 5 Dimension Health Questionnaire, SF-6D = Short Form 6 Dimensions Health Questionnaire, PCS = Pain Catastrophizing Scale, IPAQ = IPAQ, International Physical Activity Questionnaire, patient expect symptom relief = patient expectations of symptom relief, CROM = Cervical Range of Motion Device, WAD = whiplash associated disorder. X in bold indicate outcome measurements for each paper. X not in bold indicate potential factors associated with outcome measurements.

**Cost-effectiveness**

Quality adjusted life-years (QALYs) were calculated to provide a generic measurement for comparing health-related outcomes between treatments. One QALY equals one year in
perfect health, and cost-utility is expressed as cost per QALY gained. QALYS are based on HRQoL measurements, but lately disease-specific instruments such as the NDI have also been suggested to be useful (82). To be cost-effective an intervention needs to generate QALY gain, otherwise it cannot be cost-effective regardless of costs. In this thesis three different instruments were used to calculate QALYs: the EQ-5D (79) and SF-6D (143), and the NDI (48). Correlations between these instruments in acute WAD are reported to be >0.60 (SF-6D and EQ-5D), and <0.70 (NDI and SF-6D/EQ-5D) but both the EQ-5D and NDI demonstrated superior responsiveness to longitudinal health changes than the SF-6D (144). The EQ-5D was the primary HRQoL measurement in this analysis (paper 4). It contains five items describing the participant’s current health state, including mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Responses to each dimension are graded from 1 (no problem) to 3 (severe problems). In the UK, the National Institute for Health and Clinical Excellence (NICE) has specified the EQ-5D to be its preferred utility measurement, and the British value set (0.594 – 1, with 1 representing full health) was used in this thesis (79). Permission to use the EQ-5D was obtained from the EuroQol Group Foundation.

The SF-6D classification system includes the domains of physical functioning, social functioning, bodily pain, role limitation, mental health, and vitality. This tool is derived from 11 questions from the SF-36 questionnaire to be used in economic evaluations (143). Each question has between four and six response options, generating a value set from 0.296 to 1, with 1 representing full health. A licence to use the SF-36 was obtained from Quality Metrics Inc, USA.

To also use a disease specific instrument, the NDI score was converted into an SF-6D utility score as suggested by Richardson and Berven (82). Incremental cost-effective ratios (ICERs) were also calculated, as the main outcome to determine which intervention was cost-effective. ICERs represent comparisons between groups of the incremental costs associated with one additional unit of the measure of effect (QALY) and are calculated as the ratio of the cost of treatment A minus the cost of treatment B divided by the effect of treatment A minus the effect of treatment B. The main outcome for this study was the ICER from a societal perspective, including the cost of interventions, production loss, drugs and additional health care. Production loss, i.e. sick-leave, was collected from the Swedish Social Insurance Agency. Drugs and additional care were self-reported, and the intervention and additional health care costs were based on county council price lists in Sweden for 2015. The number of physiotherapist visits within the study was reported by the treating physiotherapists, who
were also asked to estimate the number of patients that could be treated simultaneously with 
the intervention in question, since it’s common in clinical practise to have more than one 
patient exercising in the gym at the same time. Benchmark ICER thresholds vary and may be 
arbitrary (145). In the UK, interventions from a health care perspective (including health care 
costs only) are often considered cost-effective when a QALY gained costs less than £30,000 
(€ 42,200) (146), whereas in other countries, e.g. Sweden, similar sums are often used from a 
societal perspective.

**Statistical analyses**

A sample size calculation was performed to avoid a type II error, i.e. a big enough sample to 
be able to detect if differences between groups exist. The required sample size was based on 
the expected difference in change for main outcome, the NDI (3.5/50, SD 7), between any of 
the three groups with an alpha level of 5% and a power goal of 80%, allowing for 10% drop- 
outs, rendering a sample-size of 216. In all 4 papers descriptive statistics are presented. If 
only one item of data was missing from the NDI scores or two items were missing from the 
SES, the missing data were substituted by the average item score of the questionnaire for that 
participant. If more data items were missing, that particular score was omitted from the 
analysis.

Types of statistics used have been presented in accordance with the guidelines of each 
paper. If the distribution of data is normal when using VAS-scales, parametric statistical 
analysis can be used according to Wiliamson and Hoggart (95) and have been treated as such. 
There are different view as to whether to treat questionnaire scores as parametric or non- 
parametric. Even though by definition they are ordinal, in the literature they are often 
statistically treated as interval data. The robust one-way ANOVA with Tukeys post-hoc 
correction was used for three-group-comparisons of parametric data, including score data in 
paper 1 and 4. Due to a small but significant difference in age and gender between groups 
these covariates were not used for an ANCOVA (147, 148). There was however no 
correlation between either age or gender and any of the outcomes. For within-group 
differences Repeated Measure ANOVA with Mauchly or Greenhouse-Geisser correction 
depending on sphericity with post-hoc Bonferroni correction was used. Data was normally 
distributed and there was homogeneity of variances. Normal distribution was determined by 
visual inspection (in cooperation with a statistician), and Kolmogorov-Smirnov test (normal
distribution >0.5) and homogeneity of variance was determined by Levene’s test (equal variance >0.05) (147). Non-parametric between-group comparisons, including score data in paper 3, were evaluated with the Kruskal-Wallis test, with the Mann-Whitney U test for post-hoc. Non-parametric within-group differences were calculated using a Friedman’s ANOVA with post-hoc Wilcoxon’s signed rank test in accordance with the guidelines of that particular journal. All score data in this thesis has been tested both with parametric and non-parametric tests without any substantial differences between results. In comparisons with binary outcomes Chi-Square tests were used. Independent samples t-tests for parametric data or Mann Whitney U-test for ordinal data were used for drop-out analyses. Pearson’s correlation test was used to assess correlation for parametric and bivariate outcomes whereas Spearman’s test was used for non-parametric outcomes.

The primary analyses were performed on an intent-to-treat basis, including all participants completing each measurement without any imputations of missing data, except in paper 3. Participants’ results were thus analysed in accordance with randomization group, regardless of adherence to that intervention. To determine the proportion of participants reaching MCID in each group, sub-analyses of participants who had reached the predefined cut-off values as described previously was also performed. Statistical significance was set at p<0.05 (with Bonferroni post-hoc correction at 0.017).

In paper 2 a multivariable logistic regression using stepwise backward regression with p ≥ 0.1 as a limit for removal of variables to reduce the risk of overlooking potential predictors was used. A sample-size of 10 subjects per variable in a regression equation is reported to be adequate (149). The linear regression technique was performed to check multicollinearity. A tolerance level <0.1 or variance inflation factor (VIF) >10 defined significant multicollinearity (147), but no variables needed to be excluded (All VIF <3.1 and tolerance levels >0.33). Goodness of fit was assessed using the Hosmer-Lemeshow test (147), and model assessment was made using Nagelkerke’s pseudo R². Group comparisons of the significant variables found in the final step of the regression were performed using the Mann-Whitney U-test, independent sample t-test or chi-square as appropriate.

In paper 3 missing scores were imputed at 12 and 24 months due to a larger number of missing data at 24 months (43% missing data), using closest match imputation, which even when missing data is high (up to 60%), is considered an efficacious and reliable imputation method in repeated measures data (150). Data was considered missing at random, i.e. missing
MATERIAL AND METHODS

independently of unobserved data, since there was for instance no difference in pain or
disability reduction at one year for those who completed the 24-month-follow-up, and those
who did not. The method replaces a participant’s missing time point with a value obtained
from another participant with similar scores on the same measurement, assessed at other time
points which are available for both donor and recipients. Possible donors were participants
with complete measurements from at least three time points, and participants with one
missing data point served as recipients. The score obtained by the donor at that time point
was imputed to the missing time point for the recipient. Donors with less than 15% absolute
differences between the recipient’s scores and donor’s scores at all other time points for
which data was present, and with the same trend over time (better, worse, or unchanged) were
defined as closest match. Absolute matches were given priority and each outcome was
analysed and imputed separately. The closest match donor was chosen at random from the
possible donors by the computer (SPSS version 22) when more than one possible donor was
identified. All analyses were also re-run without imputed values and did not cause a
significant difference in the results.

In paper 4 The NDI score was converted into an SF-6D utility score using the formula with
the highest $R^2$ value proposed by Richardson and Berven (82): - 0.0135×NDI + 0.8636.
The SPSS versions 19-23 (SPSS Inc, Chicago, IL, USA) have been used.

Ethical considerations
This study was implemented in accordance with Swedish laws and the declaration of
Helsinki. The study was approved by the Ethics Committee at the Faculty of Health Sciences
at Linköping University in Sweden (2010/1888-31 and 2011/262-32). Informed consent was
obtained from all participants who are included in the study. Exercise can be considered a
safe intervention, and there were no known risks associated with participation in the study
except temporary and benign side effects (7). Treating physiotherapists and test leaders
involved in the study were registered at the National Board of Health and Welfare in Sweden.
All personal details of the participants were anonymous at data-entry. There were no
commercial interests tied to the study.
RESULTS

Disability and functioning

Between-group results showed that on a group level, using intention-to-treat analyses, both the NSE and NSEB groups were significantly more improved than the PPA-group at all time points regarding disability (NDI). Regarding functioning (PSFS) the result was the same at 3 months (Landén Ludvigsson, previously unpublished data for 3 and 6 months, 2013). At 6 months the difference was only significant between the NSEB and PPA group, but at 12 and 24 months the difference was only significant between the NSE and PPA group. There was no significant difference between the NSE and NSEB groups however (table 5 and papers 1, 3), (101).

With regard to clinically relevant improvement in disability (NDI score, ≥5 points reduction), there was no significant difference between groups (p=0.23) at 3 months, but there was a significant difference at 6 months (p=0.02) where the NSEB group had the largest proportion of improved participants (figure 4 and paper1) (101). A greater proportion of participants in both the NSE and NSEB groups compared with the PPA group reported clinically relevant improvement in disability (NDI) at 12 months (p <0.01) and 24 months (p=0.01). The difference regarding functioning (PSFS) was statistically significant at all time points, except 24 months (p=0.32) (figures 4, 5 and paper 3), (133).

Within-group results revealed that disability levels in both the NSE and NSEB groups significantly improved over time (NDI, P<0.01/<0.001) with no significant improvement observed for the PPA group (papers 1, 3), (101, 133). The improvements were gained in the first 6 months and were maintained over time (paper 3), (133). Regarding functioning (PSFS) all three groups improved in the early stages: NSE both from baseline to 3 and 6 months, NSEB to 6 months only (both p<0.001.), and PPA to 3 months only (PPA <0.01) (Landén Ludvigsson, unpublished data 2013). The results only remained improved over time for the two neck-specific groups (NSE 0.01, NSEB < 0.001, PPA 0.13) and there was no significant further gain after 6 months (paper 3), (133).
Table 5. Change scores from baseline to 3, 6, 12 and 24 months

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>NSE</th>
<th>n</th>
<th>NSEB</th>
<th>n</th>
<th>PPA</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDI</td>
<td>62</td>
<td>-2.2 (6.0)</td>
<td>66</td>
<td>-2.7 (4.6)</td>
<td>59</td>
<td>-0.3 (4.3)</td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>grade 2</td>
<td>41</td>
<td>-2.1 (6.2)</td>
<td>29</td>
<td>-2.2 (4.8)</td>
<td>35</td>
<td>-0.6 (3.8)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>21</td>
<td>-2.4 (5.5)</td>
<td>37</td>
<td>-3.1 (4.4)</td>
<td>24</td>
<td>0 (5.1)</td>
<td></td>
</tr>
<tr>
<td>PSFS</td>
<td>61</td>
<td>1.5 (1.9)</td>
<td>66</td>
<td>1.6 (1.9)</td>
<td>56</td>
<td>0.4 (2.0)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>grade 2</td>
<td>40</td>
<td>1.5 (2.1)</td>
<td>29</td>
<td>1.3 (2)</td>
<td>33</td>
<td>0.4 (1.8)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>21</td>
<td>1.4 (1.5)</td>
<td>37</td>
<td>1.8 (1.7)</td>
<td>23</td>
<td>0.4 (2.3)</td>
<td></td>
</tr>
<tr>
<td>VAS, P</td>
<td>64</td>
<td>-10 (28)</td>
<td>67</td>
<td>-9 (24)</td>
<td>58</td>
<td>-2 (25)</td>
<td>0.17</td>
</tr>
<tr>
<td>grade 2</td>
<td>42</td>
<td>-9 (27)</td>
<td>30</td>
<td>-7 (24)</td>
<td>35</td>
<td>2 (22)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>-12 (30)</td>
<td>37</td>
<td>-11 (25)</td>
<td>23</td>
<td>-8 (29)</td>
<td></td>
</tr>
<tr>
<td>VAS B</td>
<td>64</td>
<td>-13 (28)</td>
<td>67</td>
<td>-10 (23)</td>
<td>58</td>
<td>-3 (25)</td>
<td>0.09</td>
</tr>
<tr>
<td>grade 2</td>
<td>42</td>
<td>-15 (28)</td>
<td>30</td>
<td>-10 (21)</td>
<td>35</td>
<td>2 (22)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>-12 (27)</td>
<td>37</td>
<td>-9 (25)</td>
<td>23</td>
<td>-10 (29)</td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>60</td>
<td>12 (28)</td>
<td>62</td>
<td>6 (25)</td>
<td>58</td>
<td>6 (25)</td>
<td>0.32</td>
</tr>
<tr>
<td>grade 2</td>
<td>40</td>
<td>15 (31)</td>
<td>27</td>
<td>2 (20)</td>
<td>35</td>
<td>3 (17)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>20</td>
<td>8 (22)</td>
<td>35</td>
<td>9 (27)</td>
<td>23</td>
<td>11 (34)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n</th>
<th>NSE</th>
<th>n</th>
<th>NSEB</th>
<th>n</th>
<th>PPA</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDI</td>
<td>56</td>
<td>-2.5 (6.3)</td>
<td>57</td>
<td>-3.7 (5.8)</td>
<td>53</td>
<td>-0.2 (5.7)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>grade 2</td>
<td>38</td>
<td>-2.1 (6)</td>
<td>24</td>
<td>-3.5 (6.3)</td>
<td>32</td>
<td>0.6 (4.6)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>18</td>
<td>-3.4 (6.9)</td>
<td>33</td>
<td>-3.8 (5.5)</td>
<td>21</td>
<td>-0.5 (7.2)</td>
<td></td>
</tr>
<tr>
<td>PSFS</td>
<td>52</td>
<td>1.8 (2.3)</td>
<td>58</td>
<td>2.5 (2.3)</td>
<td>51</td>
<td>0.9 (2.6)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>grade 2</td>
<td>35</td>
<td>1.9 (2.4)</td>
<td>24</td>
<td>2.4 (2)</td>
<td>31</td>
<td>0.9 (2)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>17</td>
<td>1.6 (2)</td>
<td>34</td>
<td>2.5 (2.5)</td>
<td>20</td>
<td>1 (3.4)</td>
<td></td>
</tr>
<tr>
<td>VAS, P</td>
<td>57</td>
<td>-10 (24)</td>
<td>56</td>
<td>-16 (24)</td>
<td>52</td>
<td>-8 (20)</td>
<td>0.21</td>
</tr>
<tr>
<td>grade 2</td>
<td>38</td>
<td>-8 (24)</td>
<td>23</td>
<td>-12 (26)</td>
<td>32</td>
<td>-7 (19)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>19</td>
<td>-15 (28)</td>
<td>33</td>
<td>-18 (22)</td>
<td>20</td>
<td>-9 (23)</td>
<td></td>
</tr>
<tr>
<td>VAS B</td>
<td>57</td>
<td>-15 (26)</td>
<td>56</td>
<td>-17 (24)</td>
<td>53</td>
<td>-9 (25)</td>
<td>0.20</td>
</tr>
<tr>
<td>grade 2</td>
<td>38</td>
<td>-16 (28)</td>
<td>23</td>
<td>-12 (24)</td>
<td>32</td>
<td>-7 (22)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>19</td>
<td>-12 (23)</td>
<td>33</td>
<td>-20 (23)</td>
<td>21</td>
<td>-11 (28)</td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>53</td>
<td>11 (28)</td>
<td>54</td>
<td>1 (31)</td>
<td>51</td>
<td>3 (37)</td>
<td>0.32</td>
</tr>
<tr>
<td>grade 2</td>
<td>36</td>
<td>12 (27)</td>
<td>24</td>
<td>-1 (28)</td>
<td>31</td>
<td>5 (21)</td>
<td></td>
</tr>
<tr>
<td>grade 3</td>
<td>17</td>
<td>7 (29)</td>
<td>30</td>
<td>3 (34)</td>
<td>20</td>
<td>-2 (53)</td>
<td></td>
</tr>
</tbody>
</table>

*P.T.O for the rest of the table*
RESULTS

NSE = neck-specific exercise group, NSEB = neck-specific exercise group with a behavioural approach, PPA = prescription of physical activity group, NDI = Neck Disability Index, PSFS = Pain Specific Functional Scale, VAS P = Pain Visual Analogue Scale, VAS B = Pain Bothersomeness Visual Analogue Scale, SES = Self-efficacy Scale. Post-hoc results: ¹significant between NSE and PPA only ²significant between NSEB and PPA only. P-values are based on one-way ANOVA with Tukey’s correction for post-hoc.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>12 months</th>
<th></th>
<th>24 months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>NSE</td>
<td>n</td>
<td>NSEB</td>
</tr>
<tr>
<td>NDI</td>
<td>65</td>
<td>-2.8 (7.0)</td>
<td>68</td>
<td>-3.9 (6.5)</td>
</tr>
<tr>
<td>grade 2</td>
<td>44</td>
<td>-2.3 (7.0)</td>
<td>31</td>
<td>-3.4 (6.5)</td>
</tr>
<tr>
<td>grade 3</td>
<td>21</td>
<td>-3.8 (7.1)</td>
<td>37</td>
<td>-4.4 (6.5)</td>
</tr>
<tr>
<td>PSFS</td>
<td>62</td>
<td>2.3 (2.4)</td>
<td>66</td>
<td>2.1 (2.7)</td>
</tr>
<tr>
<td>grade 2</td>
<td>40</td>
<td>2.2 (2.4)</td>
<td>30</td>
<td>2.1 (2.6)</td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>2.4 (2.4)</td>
<td>36</td>
<td>2.0 (2.9)</td>
</tr>
<tr>
<td>VAS P</td>
<td>66</td>
<td>-13 (23)</td>
<td>68</td>
<td>-12 (25)</td>
</tr>
<tr>
<td>grade 2</td>
<td>44</td>
<td>-13 (22)</td>
<td>31</td>
<td>-9 (23)</td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>-13 (26)</td>
<td>37</td>
<td>-15 (26)</td>
</tr>
<tr>
<td>VAS B</td>
<td>66</td>
<td>-15 (27)</td>
<td>68</td>
<td>-13 (26)</td>
</tr>
<tr>
<td>grade 2</td>
<td>44</td>
<td>-15 (29)</td>
<td>31</td>
<td>-12 (27)</td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>-14 (24)</td>
<td>37</td>
<td>-14 (27)</td>
</tr>
<tr>
<td>SES</td>
<td>64</td>
<td>6 (26)</td>
<td>68</td>
<td>8 (31)</td>
</tr>
<tr>
<td>grade 2</td>
<td>42</td>
<td>6 (28)</td>
<td>31</td>
<td>4 (21)</td>
</tr>
<tr>
<td>grade 3</td>
<td>22</td>
<td>5 (24)</td>
<td>37</td>
<td>11 (38)</td>
</tr>
</tbody>
</table>

| NDI     | 61        | -1.8 (6.1) | 66        | -3.7 (6.4)| 55        | 0.6 (5.7) | <0.001²  |           |           |           |           |           |           |          |
| grade 2 | 41        | -0.8 (6.1) | 31        | -3.8 (7.1)| 34        | -0.4 (7.6)|          |           |           |           |           |           |           |          |
| grade 3 | 20        | -3.2 (6.0)| 35        | -3.5 (5.9)| 21        | 2.2 (5.5) |          |           |           |           |           |           |           |          |
| PSFS    | 62        | 1.7 (2.4)  | 63        | 1.5 (2.7)| 55        | 0.5 (2.3) | 0.02¹    |           |           |           |           |           |           |          |
| grade 2 | 40        | 1.7 (2.7)  | 29        | 1.4 (2.8)| 33        | 0.6 (2.3) |          |           |           |           |           |           |           |          |
| grade 3 | 22        | 1.7 (2.0)  | 34        | 1.5 (2.5)| 22        | 0.3 (2.2) |          |           |           |           |           |           |           |          |
| VAS P   | 62        | -14 (25)   | 54        | -14 (22)| 52        | -10 (26)  | 0.67     |           |           |           |           |           |           |          |
| grade 2 | 41        | -11 (25)   | 28        | -11 (21)| 33        | -9 (24)   |          |           |           |           |           |           |           |          |
| grade 3 | 21        | -19 (24)   | 36        | -15 (23)| 19        | -12 (30)  |          |           |           |           |           |           |           |          |
| VAS B   | 59        | -12 (29)   | 65        | -14 (26)| 52        | -12 (28)  | 0.93     |           |           |           |           |           |           |          |
| grade 2 | 41        | -12 (31)   | 30        | -18 (25)| 34        | -8 (26)   |          |           |           |           |           |           |           |          |
| grade 3 | 18        | -13 (25)   | 35        | -5 (27) | 18        | -19 (30)  |          |           |           |           |           |           |           |          |
| SES     | 62        | 5 (24)     | 62        | 7 (36)  | 56        | -3 (25)   | 0.15     |           |           |           |           |           |           |          |
| grade 2 | 41        | 3 (22)     | 29        | 4 (35)  | 34        | -2 (22)   |          |           |           |           |           |           |           |          |
| grade 3 | 21        | 7 (27)     | 33        | 9 (37)  | 22        | -5 (27)   |          |           |           |           |           |           |           |          |
RESULTS

Pain

Between-group results regarding neck pain showed that on a group level, using intention-to-treat analyses, that there were no significant differences in change score between groups regarding pain at any time point (table 5 and paper 1, 3) (101, 133).

With regard to clinically relevant improvement in pain (≥50% reduction) there were however significantly more participants in the NSE/NSEB groups, compared with the PPA group (P<0.01) who reached this level at 3 months but not in current pain (P=0.39) (paper 1) (101). At 6 months there was no difference, but greater proportions of participants in the NSE/NSEB groups reported clinically relevant improvements in pain/pain bothersomeness at 12 months (p ≤ 0.02). At 24 months the difference was insignificant (p ≥ 0.11) (figures 4,5 and paper 3) (133).

The within-group analyses of current pain and pain bothersomeness revealed that both were significantly reduced over time up to 6 months in all the three groups (NSE and NSEB, P<0.001, PPA, P=0.04) (paper1) (101). The results were maintained for the NSE/NSEB groups for both current pain and pain bothersomeness up to the final follow-up at 24 months (NSE, NSEB p ≤ 0.001). Pain bothersomeness was also maintained in the PPA group (p = 0.02), although current pain was not (p = 0.07). The improvements were gained in the first 6 months without significant changes thereafter for either outcome (Table 5, and paper 3), (133).

Self-efficacy

There were no differences between groups in self-efficacy at any of the time points. The within-group results demonstrated improvement for the NSE group only over time (p = NSE 0.02, NSEB 0.07, PPA 0.86). After 6 months, there were no significant further changes in any of the groups (p > 0.12, table 5, and paper 3), (133).
Figure 4. Percentage of participants with clinically relevant improvements at the 3 and 6 month follow-up.

The percentage of participants reaching clinical important difference in the Neck Disability Index (NDI, cut-off change 5/50), Patient Specific Functional Scale (PSFS cut-off change 2.3), P-VAS = Current Pain Visual Analogue Scale and B- VAS Pain bothersomeness Visual Analogue Scale (50% reduction). NSE = neck-specific exercise group (n=52), NSEB = neck-specific exercise group with a behavioural approach (n=56), PPA = Prescription of Physical Activity group (n=29). Significant difference between groups at *P < 0.05, **P ≤0.01, ***P ≤0.001.
Figure 5. Percentage of participants with clinically relevant improvements at the 12 and 24 month follow-up.

The proportion of participants reaching clinically important difference in the Neck Disability Index (NDI, cut-off change 5/50), Patient Specific Functional Scale (PSFS cut-off change 2.3), P-VAS = Current Pain Visual Analogue Scale and B- VAS Pain bothersomeness Visual Analogue Scale (50% reduction). NSE = neck-specific exercise group (n=44), NSEB = neck-specific exercise group with a behavioural approach (n=41), PPA = Prescription of Physical Activity group (n=41). Significant difference between groups at*P =0.02, **P ≤0.01, (133).

Factors associated with disability and pain reduction

At 3 months, a total of 77/202 participants (38%) were classified as being clinically relevantly improved, reaching the cut-off for MCID regarding neck-related disability or substantial pain reduction. Out of these participants 26 (34%) reached the MCID/substantial pain reduction according to both pain and disability. The majority of the 77 participants reached the level of substantial pain reduction (n=53, 69%), either in pain reduction only, or in combination with a disability reduction. The number of participants reaching the MCID regarding disability was similar (n=50, 65%).

At 12 months about half of the participants who reached the MCID (n=79) did so both according to pain and disability reduction (n=38, 49%). The majority of the 79 participants...
reached the level of substantial pain reduction (n=67, 85%), while somewhat fewer participants reached the level of MCID regarding disability (n=51, 64%) (figure 6).

![Figure 6. Percentage of clinically relevantly improved participants at 12 months](image)

Percentage of clinically relevantly improved participants at 12 months (n=79) who attend at least one of the study physiotherapy visits, regarding neck disability reduction (>5 points, NDI = Neck Disability Index), pain reduction (>50%, VAS scale), or both disability and pain reduction.

The only factor significantly associated with a reduction in both neck pain and neck-related disability at both 3 and 12 months, was participation in the NSE group. For instance at 3 months, the NSE group had four times higher odds of achieving clinically important disability reduction compared with the PPA group (table 6). Other factors at 3 months that were significantly associated with disability reduction were higher baseline NDI score, WAD grade 3 and higher neck extensor muscle endurance (table 6).

At 12 months, group allocation was still a significant factor for disability reduction. Participants in the NSEB group had 6.4 times and the NSE group had 5.3 times increased odds of achieving a clinically relevant reduction in neck disability compared to participants in the PPA group. The other factors significantly associated with neck disability reduction were higher baseline NDI score, and absence of dizziness. Patient expectations were almost significant (p = 0.051) (table 6 and paper 2), (134).
Table 6. Factors associated with disability reduction at 3 and 12 months.

<table>
<thead>
<tr>
<th></th>
<th>3 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>Odds ratio (95% CI)</td>
</tr>
<tr>
<td>Intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSE</td>
<td>&lt;0.01</td>
<td>4.02 (1.44-11.24)</td>
</tr>
<tr>
<td>NSEB</td>
<td>0.06</td>
<td>2.69 (0.97-7.5)</td>
</tr>
<tr>
<td>Neck disability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(NDI) high</td>
<td>0.04</td>
<td>1.07 (1.00-1.13)</td>
</tr>
<tr>
<td>WAD grade 3</td>
<td>0.03</td>
<td>2.65 (1.10-6.38)</td>
</tr>
<tr>
<td>Neck muscle endurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(extensors, high)</td>
<td>&lt;0.01</td>
<td>1.01 (1.00-1.01)</td>
</tr>
</tbody>
</table>

NSE = neck-specific exercise group, NSEB = neck-specific exercise group with a behavioural approach. NSE/NSEB are reported in comparison to prescription of physical activity group. NDI = Neck Disability Index, WAD = Whiplash Associated Disorder.

Regarding neck pain reduction, group allocation was a significant factor for both physiotherapist-led neck-specific exercise groups at 3 months. The NSE group had 3.9 and the NSEB group had 3.0 times higher odds of achieving substantial pain reduction after 3 months compared to the PPA group. The other significant factors associated with pain reduction were higher baseline pain, lower NDI score and higher range of rotation motion. The only factor predictive of pain reduction at 12 months was group allocation and more specifically so being in the NSE group. The NSE group had 3.7 times higher odds of achieving pain reduction compared to participants in the PPA group (table 7 and paper 2) (134).

Comparing participants who reached MCID/substantial disability and/or pain reduction with those who did not, there were some significant differences. For both outcomes at both time points there were more participants who reached this level in the two neck-specific exercise groups (p≤0.03) No other factor was significantly different between those who reached this level of reduction and those who did not regarding disability. Regarding pain those with a substantial pain reduction at 3 months, had a larger baseline range of motion and
lower NDI scores than those who did not (paper 2) (134). No other factor than allocation was significantly different at 12 months.

Table 7. Factors associated with pain reduction at 3 and 12 months

<table>
<thead>
<tr>
<th>Intervention group</th>
<th>p-value</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.01</td>
<td>3.85 (1.39-10.69)</td>
</tr>
<tr>
<td>NSEB</td>
<td>0.04</td>
<td>2.99 (1.05-8.51)</td>
</tr>
<tr>
<td>Neck disability (NDI), low</td>
<td>0.04</td>
<td>0.92 (0.85-1.0)</td>
</tr>
<tr>
<td>Range of motion, rotation, high</td>
<td>0.01</td>
<td>1.02 (1.00-1.04)</td>
</tr>
<tr>
<td>Pain (VAS) high</td>
<td>0.02</td>
<td>1.03 (1.00-1.05)</td>
</tr>
</tbody>
</table>

NSE = neck-specific exercise group, NSEB = neck-specific exercise group with a behavioural approach. NSE/NSEB are reported in comparison to prescription of physical activity group. NDI = Neck Disability Index, rotation = full rotation from left to right, VAS = Visual Analogue Scale, current neck pain

Excluding the randomization group in the analyses (Landén Ludvigsson, unpublished data 2016), higher NDI was still a significant factor associated with disability reduction at one year (OR 1.1, CI 1.0-1.2), but also the absence of use of analgesics (OR 2.9, CI 1.1 - 8.0) was then significant (both p<0.03). For pain reduction at 12 months, lower NDI (OR 0.93, CI 0.87-0.89) was the only significant factor associated with pain reduction without inclusion of randomization groups (p=0.03).

Cost-effectiveness

At one year, greater improvements in HRQoL according to the EQ-5D was reported by the NSEB group compared to the PPA group, which reported a deterioration of health (p<0.01). This was also the case for the NDI utility change scores for both NSE and NSEB versus PPA (p= 0.001). The difference between groups in the SF-6D utility change score did however not
reach significance (p=0.07). The improvements in health utility scores were between 0.040-0.054 (NSE), 0.028-0.106 (NSEB) and -0.038 - 0.002 (PPA). Since improvement in health is a presumption for any intervention to be cost-effective the PPA intervention could therefore not be cost-effective regardless of costs associated with the intervention. There was no significant difference in any of the HRQoL outcomes between the NSE and NSEB groups (paper 4), (151). QALY gains are presented in paper 4 (151).

The NSE group had the lowest cost of production loss (sick-leave) for 12 months post-inclusion. The NSE group also had a lower intervention cost than the NSEB group, due to the shorter duration of each session (NSE 30 min, NSEB 40 min). The mean number of visits within the study interventions were; NSE 17.5 (SD 8.5), NSEB 18 (SD 8.5), and PPA 1.5 (SD 1.5). The treating physiotherapists estimated that the mean number of patients that could be treated simultaneously in the gym was 2.5 (SD1, NSE) or 2 (SD1, NSEB, range both groups 1-5). The total costs for the interventions based on the number of patients that could be treated simultaneously/session according to treating physiotherapists were thus € 446 for NSE, € 764 for NSEB, and € 191 for PPA. Additional health care costs were significantly higher in the PPA group (p=0.04), mainly due to multimodal treatment at pain clinics. Health care costs are further presented in paper 4 (151). Fewer participants in the two neck-specific groups reported using analgesics at 12 months (paper 3), (133).

Calculating cost-effectiveness was, as mentioned, pointless for the PPA group, there was no QALY gain. The NSEB group tended to be more effective than the NSE group using the EQ-5D or NDI but, mainly due to production loss, also twice as expensive from a societal perspective. The ICERs from a societal perspective for adding NSEB vs. NSE alone were € 87,169 based on the EQ-5D or € 326,882 based on the NDI. If using the SF-6D, the NSE group dominated the NSEB group, as NSE alone tended to be both more effective and less costly than the addition of NSEB. From a health care perspective, the ICERs for adding NSEB vs. NSE were € 12,720 (EQ-5D) and € 47,699 (NDI). Depending on which HRQoL measurement was used, the intervention cost per QALY gained as opposed to no treatment at all was € 16,484 to 22,254 for the NSE group and € 14,384 to 54,453 for the NSEB group. Regardless of measurement, the NSE intervention was the most cost-effective intervention from a societal perspective since the additional costs for adding a behavioural approach per QALY gained (i.e. ICERs) were substantially higher than the NICE threshold defining cost-effectiveness (146). However this threshold can also be used from a health care perspective, and if so the ICER for adding NSEB could be considered reasonable if using the EQ-5D, but
not if using the other two HRQoL instruments. Tables of HRQoL/QALYs, health care costs and ICERs are available in paper 4, (151).
DISCUSSION

Results – main findings

Considering that little or no change can be expected after 3 months with WAD (9, 46), the results of the neck-specific exercise both with and without a behavioural approach in this thesis are promising. A substantial number of participants with chronic WAD (mean duration 20 months) obtained long-lasting symptom reduction, particularly following neck-specific exercise. Improvements were found both including all participants with intention-to-treat analyses, regardless of adherence, and in subanalyses including only adherent participants, even though the study was not powered for these subanalyses. This indicates solid results, since they were evident even in such smaller subgroup analyses.

Disability/functioning, pain and self-efficacy

Both the NSE and NSEB groups were significantly more improved than the PPA-group at all time points regarding disability (NDI) which was not significantly improved at all in the PPA-group. Both or either neck-specific group were also significantly more improved regarding functioning (PSFS) at all time points compared to the PPA group in the main intention-to-treat analyses. The PPA group was only improved regarding short-term functioning and the results did not remain over time. This supports the hypothesis that neck-specific exercise would have a better effect on disability than prescription of physical activity. Looking only at adherent participants the trend was the same, even though the differences were not significant at all time points, most likely due to insignificant power. At 6 months there was a clear trend for the NSEB group having more participants reaching the MCID than the NSE group. This may be due to this group being encouraged not to focus on temporary increase in pain, which may enhance the ability to perform various daily activities. There was no such trend at the other time-points however.

Regarding current neck pain and pain bothersomeness the difference between groups was less evident. There were no significant differences between groups on a whole group level with the intention-to-treat analyses. However regarding clinically relevant/substantial improvement in current pain/pain bothersomeness, there were significantly more participants in the NSE and NSEB groups, compared with the PPA group who reached this level at 3 and
12 months. There was a trend for more participants reaching the substantial improvement level in the NSE group compared to the NSEB group at 3 and 12 months. This may be due to the fact that exercise was supposed not to be pain provoking in the NSE group, whereas temporary increase in pain was allowed in the NSEB group.

These findings are in contrast with those of Michaleff et al. who observed similar improvements in pain and neck disability (NDI) after an extensive physiotherapy program, compared with advice in chronic WAD (75). However the exercise approaches were not the same in the study by Michaleff et al. as in this thesis. Also in their study, participants with grade 1-2 were included, not grade 2-3 as in this thesis. It is not unreasonable to assume that those with grade 1 (without any clinical neck findings) could do just as well on advice as an extensive exercise program, since there are no clinical findings to work with as a physiotherapist. They did not report the proportion of participants with grade 1 however. As presented in this thesis, among participants with WAD grade 3, disability improvements tended to be greater in the neck-specific groups, while there was rather a trend toward deterioration over time in the PPA group. WAD grade 3 has also been associated with improvement in disability (paper 2), (134). The findings are also partly different from what’s been reported by Stewart et al, who did include grade 3, but observed only short-term differences between groups receiving advice or exercise (74). However they did not specify the exercises further or clarify whether the exercises were neck-specific or not.

Regarding self-efficacy there was an overall low or non-existent effect, which was only significantly improved in the NSE group. It was surprising that it was only improved in this group and not in the NSEB group where there was a more set aim to improve self-efficacy through different strategies. This may have been due to sufficient “mastery experience” of the exercise alone, since it is reported to be the strongest influence on self-efficacy belief (152). After completing an exercise program, self-efficacy has previously been reported to improve (51), however there was a large variance in this thesis that could not be explained by level of adherence. Baseline SES score or change in SES score did not correlate with adherence (paper 1) (101), and a subgroup analysis excluding those with <50% adherence did not alter this relationship, even though mean improvements were somewhat higher in all groups in this subanalysis (Landén Ludvigsson, unpublished data 2016). Although interesting it should be acknowledged that the difference between groups was not significant. Furthermore behavioural approaches vary and other approaches may have yielded different results.
Another reason might be the relatively high baseline level of self-efficacy in this study, indicating that lack of self-efficacy was not a major problem among these participants. In the WAD classification system proposed by Söderlund and Denison, the mean level of self-efficacy in this study sample equals what in their classification can be seen in “adaptive copers”, i.e. those with the least self-efficacy problems (19). A reason for this may be that participants were recruited mainly from primary care, not from pain clinics. Among patients in pain clinics, higher levels of functional impairment and psychosocial difficulties are reported, but they may on the other hand be less representative of people in general who suffer from chronic pain (153). Looking at other psychosocial scores such as the TSK-11 (22 (SD6)) and PCS (18 (SD9)) scores in the study sample in this thesis, they were similar to previously presented scores in other chronic pain conditions (134, 139, 154).

There was no significant difference between the NSE and NSEB groups in any of the outcomes, and the hypothesis that NSEB would be better than NSE alone was therefore not confirmed. However the study was powered for difference between any of the three groups, not all three, and although speculative it is possible that a larger study sample may have yielded some differences between these groups as well. The neck-specific exercise was however the main intervention for both groups. The exercises were identical even though the approach was somewhat different. The behavioural approach was basic and may not have made a big enough difference between these groups, or the addition of a behavioural approach may not have been essential to a number of participants in this study sample. Nonetheless it is important to acknowledge that it may be important on an individual level.

Factors associated with disability and pain reduction

The only significant factor associated with a reduction in both neck related disability and neck pain at both 3 and 12 months, was participation in the NSE group. The hypothesis that that type of exercise approach would be an important factor associated with treatment response was thus supported. This also suggests that baseline factors are not as important as the actual exercise intervention. The odds were up to 5.3 times higher of achieving disability reduction and 3.9 times higher of achieving pain reduction for those in the NSE group compared to those in the PPA group. The fact that exercise was supposed to be non-pain provoking in the NSE group may have contributed to the higher odds of pain reduction in this group. Participation in the NSEB group was however also associated with pain reduction at 3
months (odds ratio, 3.0) but not at 12 months. The odds of disability reduction at 12 months were highest for the NSEB group with an odds ratio of 6.4 compared to the PPA group, even though the odds at 3 months were insignificant. This may be due to the different approach encouraging participants to perform activities despite a temporary increase in their pain, and it may have taken longer than 3 months for this approach to give effect. None of the psychosocial factors at baseline had an impact on the results (paper 2), (134). These findings are consistent with a qualitative study of people with chronic WAD undergoing rehabilitation (91). In that study baseline pain and disability, not psychosocial factors, were the only factors associated with pain and disability post treatment. This was also the case in another exercise study of people with chronic WAD undergoing exercise interventions, where the degree of psychological distress did not modify the effect of treatment in chronic WAD (75). It should be noted that to maximize the clinical applicability of the analyses in this thesis, only outcome measures which are possible for clinicians to implement in everyday practice were included. Therefore it cannot be ruled out that there may be other factors, not measured in this study, associated with treatment outcome as well. This analysis also does not include other mediators (that identify possible mechanisms through which a treatment might achieve its effects) or moderators (for whom or under what conditions the treatment works) (155). The effective interventions were both based on neck-specific exercise and improvement in neck muscle endurance may be a mediator. It was thus surprising that baseline neck muscle endurance was not more strongly associated with the outcomes. However, baseline endurance time was considerably lower than what has been documented in healthy population (156) indicating that there was the potential for substantial improvement for most participants. To analyze neck muscle endurance was not part of the scope of this thesis, but has been analysed elsewhere. Compared with participants in the PPA group, participants in the NSE and NSEB groups exhibited greater gains in dorsal neck muscle endurance (both men and women) and ventral neck muscle endurance (men only), and reported less pain provocation from the neck muscle endurance test than the PPA group (156).

The fact that a higher NDI score was associated with neck disability reduction, whereas a lower NDI score was associated with neck pain reduction after 3 months illustrates the dilemma with the use of different outcomes and time points across different prediction studies, and may partially explain the inconsistency of observed predictive factors between studies (47). Furthermore, predictive factors for symptom resolution may not be the same in a population with chronic pain as in the acute phase (157).
**Cost-effectiveness**

This is the first study to explore cost-effectiveness in the treatment of chronic WAD. Even though a very cheap intervention, PPA was not cost-effective due to no QALY gain, and could therefore not be recommended from a cost-effectiveness perspective. Whether NSE or the higher costs associated with NSEB can be recommended depends on perspective and HRQoL instrument used. From a societal perspective NSEB is not cost-effective, mainly due to higher production loss, and NSE is the cost-effective option. From a health care perspective the additional costs of adding the behavioural approach could be considered cost-effective, however only if using EQ-5D as the HRQoL instrument of choice. If using SF-6D or NDI the NSE intervention was the cost-effective option also from a health-care perspective. The results support the hypothesis that neck-specific exercise is cost-effective compared to prescription of physical activity. The results also enlighten that different HRQol measurements may generate somewhat different results, which is important to consider when comparing the results of different cost-effectiveness studies.

**WAD grade 2 and grade 3**

This study is the first study to report participants with WAD grade 2 and 3 separately. Even though the study was not powered for subanalysing these two groups it is important to report the results for these groups separately, since those with grade 3 are rarely included in other treatment studies of chronic WAD. The findings in this RCT suggest that those with grade 3 seem to benefit no less than those with grade 2 from neck-specific exercise with or without a behavioural approach. Furthermore, participants classified as WAD grade 3 had more baseline pain and disability, and our findings are in accordance with previous findings that show supervised exercise to be more effective for those with higher baseline neck pain or disability (74). After 3 months WAD grade 3 was even a significant factor associated with improvement in disability, as opposed to grade 2 (134). Whether the neurological symptoms also improved is yet to be determined. There were generally a trend for lower improvements for those with grade 3 in the PPA group regarding disability/functioning.
Methodological considerations

Study Design

The reason for choosing an RCT is to be able to compare the three interventions in a way as appropriate as possible, to make possible confounding factors as evenly distributed as possible between the groups, and to eliminate any influence on allocation from investigators, participants or clinicians. It is an effectiveness study which aims to evaluate effects in heterogeneous clinical settings, and is thus more generalizable than efficacy studies which are characterized by strong control in a standardized program, for instance in a lab, and often to very a homogeneous population (158). Having three arms requires more participants to ensure sufficient power. Even though the drop-out rate was somewhat higher than assumed in the sample-size calculation, the results demonstrate that there have been significant differences between groups in many outcomes. A limitation is the large loss to follow-up at 24 months. This may have been due to the fact that participants at this point were asked to fill out questionnaires at home and mail them back, albeit in prepaid envelopes, whereas on all previous occasions participants also underwent clinical examinations where the questionnaires were collected.

A multi-centre trial was necessary to get enough participants in time. It makes the results more generalizable, but offers less control of the performed interventions. However to minimize this impact, the treating physiotherapists received a standardized one day education regarding WAD and the study interventions before seeing any participants. Even though the treating physiotherapists were used to seeing neck pain patients, one day of training may not have been enough to optimally master the neck-specific exercise for those who had no previous experience of this kind of exercise. We tried to enhance the knowledge by providing both oral and written information, by giving the physiotherapists time to practise for at least a month before seeing any study participant, and the physiotherapists were welcome to contact the project leaders for more guidance if needed all throughout the study. All physiotherapists who took care of participants in the NSEB group had an interest, some previous knowledge and experience of behavioural approaches before involvement in the study. However it should be acknowledged that results may have been different if all physiotherapists were specialized in more advanced behavioural approaches, or if a different behavioural approach was utilized.
Participants in the PPA group only had one to two physiotherapist visits. This may have influenced the results, as the other two groups had regular contact with their physiotherapist. All participants however, also visited blinded experienced physiotherapists (project leaders) four times for other tests, and each participant was then encouraged to keep exercising in accordance with the guidelines given by their treating physiotherapist. Expectations of the three interventions were similar at baseline and regarding fulfilment of expectations at six months there was no difference between the three groups (76 to 85 on a 100 mm VAS scale, P=0.64) (paper 1), (101). Furthermore, clinician treatment preferences have not shown any moderation of treatment effect in chronic WAD (159) and as much as possible, the physiotherapists were selected and matched to work within their field of knowledge and interest.

There was no control group without any intervention in this study, which is a limitation. However, there were no significant improvement observed for the PPA group regarding the main outcome, NDI, or for instance general health. The PPA group may thus be seen as a control group. Furthermore, when comparing NSE for three months with being on a waiting list in a subgroup of this study sample (n=41, not part of this thesis), participants who took part in the NSE group were significantly improved both regarding NDI, SES, and EQ-5D compared with those on the waiting-list (76).

**Classification of WAD**

The QTF classification of WAD does not take psychosocial aspects into account and the classification system and especially its prognostic value has thus been criticized by some (17, 19). However, with a growing body of evidence of physical findings in WAD, a classification system that on the contrary does not take physical findings into account, like the one proposed by Söderlund and Denison, may thus not be more useful. Sterling suggested a combined classification system as a further development of the QTF classification system, but it requires several tests and equipment that may not be available to all clinicians, such as pressure algometry to test pain pressure thresholds, thermorollers to test thermal sensibility or EMG to test motor function (17). This may be why the QTF classification system still is used, even by these authors. The QTF classification is also similar to the International Statistical Classification of Diseases and Related Health Problems (ICD) proposed by the WHO, used in everyday clinical practice, based on clinical findings for spinal diagnoses around the world.
Diagnoses include for instance cervicalgia or cervikobrachial syndrome and lumbago with or without sciatica (160), which are similar classifications to WAD grade 2 or 3.

**Clinical relevance and adherence**

The IMMPACT concludes that it is impossible to provide specific guidelines for determining whether a group difference is clinically meaningful or not, and that is has to be decided on a case-by-case basis (131). To use a cut-off of MCID to define a responder to treatment is a way of reflecting changes in a clinical intervention that are meaningful for the patient, as opposed to just statistical significance. Central tendency analysis may not give enough information to be clinically meaningful (161). Furthermore, criteria for clinically important changes in individuals cannot be extrapolated to the evaluation of group differences (131).

However, there are different ways of defining the level of MCID, and there are limitations. First of all there are several different ways to calculate the MCID, and there is no standard method. The calculations usually involve a retrospective judgment of improvement where recall bias cannot be ruled out. There may also be different views on what improvement means, where different expectations, current health status, age, socioeconomic status, or education may play a role (162). The results can also be different depending on the chosen cut-off score. The MCID of the NDI-score was set at a cut-off score of at least 5/50 reduction defining a positive responder regarding disability in this study. This is a level previously used and recommended in WAD (114), yet it may not seem very high. However in this study sample, where the mean baseline NDI score was 16.7 (SD 6.6), a 5 point reduction equals an average of 30% improvement in disability, which can be considered quite substantial. If analyzing only those with MCID regarding NDI at 3 months, their average baseline NDI score was 28 (SD 6), and their average score reduction was 8 (SD 3), which equals an average disability reduction of 29% (Landén Ludvigsson, unpublished data 2016). Of those who reached MCID/substantial pain reduction at 3 months the average baseline VAS neck pain was 44 mm (SD 24), and their average reduction was 35 mm (SD 21) (Landén Ludvigsson, unpublished data 2016). This means that the average neck pain reduction among these participants was even as high as 80%, which ascertains that there truly was a very substantial improvement for these participants.

The classification of adherent, versus non-adherent participants varies between studies and was for the secondary analyses in this thesis based mainly on clinical reasoning between the
Regarding the main outcome, the NDI, the improvement was significantly higher among adherent participants when the intervention was completed (change score adherent 6 (SD 10), non-adherent 1.8 (SD 9) p<0.02)). If analyzing each randomization group separately the trend was the same for the two neck-specific groups (NSE adherent 6.8 (SD 11.5) non-adherent 2.8 (SD 6.5), NSEB adherent 8.2 (SD 8.6) non-adherent 3 (SD 8.8)). This can be one reason for the relatively small changes in the intention-to-treat analyses, and supports the secondary analyses taking adherence into account. For the PPA group the change score was very low regardless of adherence (adherent 0.5 (SD 8.5) non-adherent 0.2 (SD 9.9)) (Landén Ludvigsson, unpublished data 2016). Only including adherent participants generates a smaller sample than what was suggested from the sample size calculation. However, whether all participants were included in the analysis, regardless of adherence, or not, the results were the same, except the improvements tended to be lower when non-adherent participants were included. It should however be acknowledged that the level of adherence was self-reported after 3 months in the NSE/NSEB groups, and at all time points in the PPA group, and there may be discrepancies between reported and actual adherence. At 3 months the PPA group had a lower level of adherence which may have been partly due to a greater amount of missing exercise data for this group at this time point. At later time points, when data was self-reported for all groups adherence was similar except at 24 months when the PPA group reported better adherence than the NSEB group. Yet the NSEB group still reported less disability. One reason might be the included discussion of strategies to handle relapse/periods of worsening in the NSEB group. It may also indicate that choice of exercise intervention is important even after 2 years, even when continuous adherence is not optimal, which has also been reported in chronic neck pain by Ylinen et al. Improvements achieved through long-term training were maintained over time despite faltering adherence (163).

**Analyses of those who declined participation and drop-outs**

The number of potential participants that were not eligible may at the first glance look very large, but of those with a cervical, but without WAD-diagnoses at the time of the health care visit, only a small proportion could be expected to have a whiplash trauma as the cause of their problems. A total of 4166 letters were sent to this group without WAD-diagnoses rendering only 30 of the 216 participants who were recruited to the study. Also, only 30% responded to the letters in this group, and a larger proportion of interested participants were excluded from this group since a whiplash trauma as the cause of their symptoms could not
be verified. Regarding those with whiplash-diagnoses (n=3784), assuming 50\% of those having sought care due to acute whiplash associated disorders would have no remaining WAD (4), that reduces the number of potential participants by half. Also there was no way of knowing whether the registered health care visit was because of an acute whiplash trauma or already chronic WAD, and most likely a fair number of individuals may already have had WAD for longer than the 36-month limit for inclusion in this study at the time of the registered health-care visit. Among those with a whiplash-diagnosis there was a higher proportion who responded to the invitation letters (55\%).

Age, gender, NDI and VAS was collected also for people with WAD-diagnoses who claimed to fulfil the self-reported criteria, but declined participation. Analysis of those who fulfilled the criteria, but declined to participate in this study, showed that the study sample was well representative in gender, age and level of pain (declined to participate, gender female 62\%, mean age 39 (SD 17), VAS 45 (SD 24)), but a higher level of disability (NDI 32 (SD 17)) was reported by those who declined participation than the study sample. Since those with more disability previously have been reported to benefit more (74) in WAD it is however not unlikely that the results would be at least equally good in the whole population. Nonetheless, whether those who declined participation would have been willing to participate in an exercise intervention is unknown. It is possible that there is thus a bias in the willingness to exercise between the study sample and those who declined participation. There were more women than men in this study, but this is consistent with the general WAD population (18).

There was no difference in the baseline outcomes (NDI, PSFS, VAS or SES), or baseline variables (age, gender, WAD grade, or allocation) (p>0.08) between participants who dropped out and those who completed for the first 12 months, except drop-outs at 12 months were however somewhat younger (age 37 (SD 11) years) than those who completed the study (age 41 (SD 11) years, p = 0.04). Since age was not correlated with the outcomes there is no indication that the there is a bias in the results regarding drop-outs up to 12 months. The drop-outs at 24 months reported somewhat more baseline disability than those who completed (133), but otherwise there were no differences. Since higher baseline NDI was a factor associated with clinically relevant disability reduction in this sample (paper 2), (134) this may indicate that the presented results at 24 months may be too conservative, and that improvements in disability may have been higher.
Outcome measurements

When clinical trials involve conditions in which there is no objective outcome measurement of improvement, such as morbidity or biomarkers, which is the case in WAD, patient reported outcomes can be used as primary outcome measures. An advantage is that they are directly reported by the individual without interpretation of the response by anyone else, and they provide information about the personal experience, as opposed to clinical findings (164). The NDI was therefore chosen as the primary outcome. The NDI is valid and reliable, and a commonly used self-reported instrument to evaluate neck-related disability (48). Both significant and clinically relevant improvements and between-group changes have been reported in this study. If analysing the baseline two-factor structure suggested (117) both seem to contribute in a similar way to the total score (“pain and interference with cognitive function” (14 (SD 4)), and “functional disability” (12 (SD 3)), (Landén Ludvigsson, unpublished data 2016).

Consistent with our findings, the PSFS is reportedly more sensitive to change than the NDI (122). This difference in sensitivity might be because the chosen activities are based on what is most important to each individual. The PSFS is constructed for use on an individual level, and has been found reliable as a clinical tool of evaluation (165). Compared to fixed-item tools there are advantages, since fixed-item tools may be difficult to interpret on an individual patient level. These tools do not consider patients’ preferences and variability in performance on particular activities (121). Using the PSFS the activity can be more specific and quantified and thus can be a valuable complement to the NDI. But that also makes it difficult to compare the outcome between different individuals. Still it is commonly used in research and group comparisons are made (74, 166, 167), and it has been shown to correlate to a great extent with the NDI (112). Changes of functioning regarding meaningful individual activities for each participant are thus what the result should be interpreted as.

The VAS scale is widely used and recommended, but there are weaknesses of the instrument including the lack of experimental grounding for the maximal anchor. “No pain” might not be that difficult to understand, whereas worst imaginable pain might be hypothetical, and depends on the individual’s experience to date. This might affect repeated individual measures since the experience of pain might have altered since the previous measurement. Another difficulty is the concept of the experience of pain itself, which has many components. To some people the scale might represent pain intensity, but to others it
could represent the emotional component of their pain (103). It is therefore recommended to be used together with other instruments which reflect the multidimensionality of the construct of interest, in this case pain (103, 123). In this thesis both pain and pain bothersomeness are therefore used as outcomes, and pain is also considered in both the NDI, PSFS and SES.

Even though frequently used in WAD-studies and found reliable, the SES does not seem to be validated in this context (168). Also some of the questions might be less relevant for some people, like for instance working on a house repair, raking leaves and shoveling snow. Bunketorp also suggests that the sum score is characterized by random disagreement, which limits the instrument’s potential to identify true treatment effects, i.e. responsiveness. Random disagreements tend to encourage Type II errors, increasing the likelihood of accepting the null-hypothesis. She suggests that the use of a median score potentially improves the reliability by lack of disagreement (129). However whether median or mean scores were used in this thesis, the results did not significantly change.

Cost-effectiveness analyses

There are some limitations regarding the data used in the cost-effectiveness analyses. Apart from production loss, and number of visits within each intervention, data was self-reported. Individual register data on additional publically funded health care consumption due to neck pain was only available in one region due to different regional interpretations of the law (43% of participants, physician and physiotherapist visits only). The second and third largest regions (42% of participants) provided useful data for physicians at hospitals only, and this data was on an allocation group level, not individual level. Register data from those participants where physician visits was available was compared with self-reported data on the number of health care visits due to neck pain in these participants. There was no difference between self-reported and register data (p <0.21). Self-reported data was therefore used. However there was a significant difference in physiotherapist visits, but regarding the number of visits within the study intervention these were reported by the treating physiotherapists. Production loss due to time off work to participate in the interventions was not registered, and it is unknown how many participants needed to take time off.

Even though the standard procedure is to use general HRQoL instruments in cost-effectiveness calculations, it can be debated. For example, the EQ-5D takes gait ability into consideration, which is less relevant to an individual with neck pain, but there is no
equivalent question about neck-related disability. In theory, this could generate different scores in similarly disabling conditions just based on anatomical location. Also, the utility scores in both EQ-5D and SF-6D are derived from the public, which may interpret health state descriptions differently than those who actually suffer from a certain condition. For example, the public may be unaware of aspects of unfamiliar health states and how people adapt to their illness (169).

**Statistics**

The sample-size calculation, based on the main outcome NDI, allowed for a 10% drop-out rate. Even though acceptable for the first 12 months, the drop-out rate was larger. However, there were significant differences between groups at all time points regarding NDI, suggesting that the power was sufficient to detect differences even in a somewhat smaller sample. This may not have been the case regarding pain however. There was a trend for better results regarding pain at all time points for the two physiotherapist-led neck-specific exercise groups compared to the PPA group, but the differences were not significant. This may be due to insignificant power, since the power calculation was based on the NDI, or due to a number of participants not being adherent to the allocated interventions. The intention-to-treat analyses, including all participants according to allocation, regardless of completion of intervention, is often recommended (170) since it reduces possible bias if disrupting the baseline equivalence established by the randomization process. However when including non-adherent participants whose results are not as good, it may affect the results in a more conservative way. In the subanalyses including only adherent participants, the proportion of participants in the NSE/NSEB groups were significantly higher than the PPA group at 3 and 12 months also regarding pain. This indicates solid results for these results even in a smaller sample, but cannot rule out insufficient power to detect differences between groups in the subanalyses of those outcomes/at those time points without significant changes.

The percentage of missing data up to 12 months was considered acceptable, but at 24 months was substantial (43% drop-out rate). Imputation methods produce a conservative estimate of treatment effect, and no imputation method can give a completely unbiased estimate of the treatment effect (170). However the closest match imputation method, which was used, is considered a reliable and efficacious imputation method in repeated measures data, even when
missing data is substantial (150) and was thus chosen for the long-term follow up. Yet the long-term results should thus be interpreted with some caution.

At baseline there was a small, but significant, difference between randomization groups regarding gender and age; however, neither factor was associated with outcomes in this study sample (101, 133) and is thus unlikely to have affected the results.
CONCLUSIONS

In this thesis physiotherapist-led neck-specific exercise with or without a behavioural approach resulted in better outcomes compared to prescription of physical activity regarding disability and functioning, as well as proportion of participants with clinically relevant improvements in these outcomes and pain. The results were gained in the first 6 months, and remained more improved over time for the two physiotherapist-led neck-specific exercise groups compared with the prescription of physical activity group.

The observed benefits of adding a behavioural approach to neck-specific exercise in this thesis were however inconclusive, and from a societal perspective, neck-specific exercise alone is the cost-effective option. Factors associated with clinically relevant improvements following these exercise interventions in chronic WAD differed whether disability or neck pain was the outcome, but also differed in the short and long term. Participation in the neck-specific exercise group without a behavioural approach (NSE) was the only factor associated with both outcomes after both 3 and 12 months, and was thus more important than the baseline factors.
The results of this thesis indicate that a considerable number of individuals with chronic WAD (in this thesis with a mean duration of 20 months) can achieve long-lasting symptom reduction, particularly following neck-specific exercise. Participants with WAD grade 3 (with neurological signs), are rarely included in other studies, but the results in this thesis support that people with both WAD grade 2 and 3 can tolerate and benefit from neck-specific exercise. The addition of the behavioural approach was inconclusive, and there were no significant differences in outcomes between the two neck-specific groups. This does not mean that the behavioural approach was not an important addition to some participants. However neck-specific exercise alone (NSE) was the only intervention associated with both disability and pain reduction both in the short and long term, and it was also the cost-effective treatment option from a societal perspective. Prescription of physical activity, though cheaper, was not cost-effective due to insufficient improvement of health, resulting in no QALY gain in this group. It is however important to consider that even so, there were individuals in all intervention groups that were clinically relevantly improved, and that there may be individual factors and circumstances that should be considered before deciding which exercise treatment is best for each individual with chronic WAD.
FUTURE RESEARCH

The chronic WAD population is heterogeneous, and it is unlikely that there is an intervention that fits all individuals, even though the results of this thesis indicate that neck-specific exercise on a group level is more effective than prescription of physical activity. It is thus important to further investigate which patients will benefit more from which kind of exercise approach. The results of the papers in this thesis also need to be confirmed in other studies. Since most studies don’t include participants with WAD grade 3, treatments for this group remain a larger gap of knowledge and it is important for future studies to also include these individuals, like in this thesis. Furthermore it is important to examine whether those with grade 2 and 3 may need different treatment approaches. Finally it is also unknown if the neurological deficits in those with grade 3 improve with neck-specific exercise, which also needs to be examined in future studies.
**Bakgrund:** Av dem som drabbas av whiplashvåld mot nacken visar internationell forskning att ca 50% fortsätter att ha nackbesvär även 1 år efter skadetillfället. Det innebär både personligt lidande och stora kostnader för samhället. Det finns både fysiska och psykosociala faktorer som kan påverka att besvären kvarstår, men forskningen har ännu inte kunnat ge svar på hur behandling av whiplashrelaterade besvär (WAD) framgångsrikt kan behandlas. Alltmer forskning visar dock på att funktionen och samspelet mellan djupa och ytligare nackmuskler förändras hos personer med WAD.

**Syfte:** Det övergripande syftet med denna avhandling var att utvärdera effekten gällande självsattad nackfunktion, smärta och tilltro till sin förmåga att kunna utföra aktiviteter trots smärta, efter tre olika träningsbehandlingar hos personer med långvarig WAD, samt att utvärdera kostnadseffektiviteten av dessa.

**Material och metod:** Totalt 216 personer med långvarig WAD sedan minst 6, max 36 månader ingick i denna randomiserade, kontrollerade studie. De var klassade som WAD grad 2 (med nacksmärta och lokala undersökningsfynd i nacken) eller grad 3 (som grad 2, men även neurologiska besvär från nacken som strålar ut i armen). Deltagarna lottades till en av tre träningsbehandlingar som pågick i 12 veckor: nackspecifik träning hos sjukgymnast, nackspecifik träning hos sjukgymnast med ett beteendemedicinsk tillägg där olika smärthanteringsstrategier ingick, eller förskrivning av fysisk aktivitet på recept att utföras utanför sjukvårdens ramar. Utvärderingen gjordes efter 3, 6, 12 och 24 månader med följande frågeformulär: Neck Disability Index (som mäter funktionsnedsättning i nacken), Patient-specifik funktionell skala (som mäter nackfunktion på egenvalda aktiviteter), Visuell Analog Skala (VAS, som mäter smärta), samt Self-Efficacy Scale (som mäter tilltro till egen förmåga att kunna utföra aktiviteter trots smärta). Analyser gjordes också av vilka faktorer som har betydelse för att smärta och funktionsnedsättning ska förbättras på ett sätt som upplevs viktigt för deltagaren, samt av den hälsoekonomiska kostnadseffektiviteten. Kostnadseffektivitet beräknas genom att jämföra kostnaden för att uppnå förbättrat hälsa mellan olika behandlingar, och mätt ur ett samhällsperspektiv inkluderar detta kostnader både för behandlingen, annan vård och sjukvårdsinkostnader.

**Resultat:** Vad gäller funktion och funktionsnedsättning var förbättringarna störst i de två sjukgymnastledda träningsgrupperna, innefattande nackspecifik träning, vid alla mätpunkter. Förbättringarna uppnåddes under de första 6 månaderna, och kvarstod sedan under resterande
mätningar. Den grupp som fick förskrivning av fysisk aktivitet på recept förbättrades inte alls avseende funktionsnedsättningen, men de rapporterade viss kortvarig förbättring av funktion av egenvälda aktiviteter. Sätt till andelen deltagare som uppnått en nivå av förbättring som anses viktig var detta också större i de två sjukgymnastledda grupperna med nackspecifik träning.

Vad gäller smärta förbättrades alla tre grupperna, och det fans ingen säker gruppsskillnad. Satt till andelen deltagare som hade minskat sin smärta med minst 50 % (= viktig förbättring) var det dock fler i de två sjukgymnastledda grupperna med nackspecifik träning vid 3 och 12 månaders-mätningarna. Vad gäller tilltro till egen förmåga att kunna utföra aktiviteter trots smärta var det bara gruppen med rent nackspecifik träning, utan beteendemedicinskt tillägg; som förbättrades men skillnaden mellan grupperna var inte signifikant. Det fanns ingen signifikant skillnad mellan de två sjukgymnastledda grupperna med nackspecifik träning gällande något mått.

Deltagande i den rent gruppen med rent nackspecifik träning, utan beteendemedicinskt tillägg, var den enda faktor som hade betydelse för att uppnå den nivå av förbättring som anses viktig både gällande funktionsnedsättning och smärta, både på kort sikt (3 månader) och längre sikt (12 månader). Oddsens var då upp till 5,3 gånger högre än för dem som fått fysisk aktivitet på recept. Övriga faktorer av betydelse skilde sig beroende på tidpunkt och om det var smärta eller funktionsnedsättning som var det som utvärderades.

Vad gäller kostnadseffektivitet uppnåddes ingen hälsoförbättring för dem som fått fysisk aktivitet på recept, och denna behandling var därför inte kostnadseffektiv trots den lägre behandlingskostnaden. Av de båda sjukgymnastledda träningsgrupperna var den rent nackspecifika gruppen utan beteendemedicinskt tillägg det mest kostnadseffektiva alternativet från ett samhällsperspektiv.

Slutsats: Sjukgymnastled nackspecifik träning resulterade i större förbättringar av funktion/funktionsnedsättning och hälsa, samt en större andel deltagare som uppnådde en viktig förbättring av såväl funktion/funktionsnedsättning och smärta, jämfört med förskrivning av fysisk aktivitet på recept. Effekten av det beteendemedicinska tillägget gav inga signifikanta ytterligare förbättringar, men kan vara av vikt för vissa individer. Att delta i den sjukgymnastledda gruppen med rent nackspecifik träning var den enda faktor av betydelse för att uppnå vad som klassas som en viktig förbättring både vad gäller smärta och funktionsnedsättning både på kort och längre sikt och den var också mest kostnadseffektiv.
Övriga faktorer av betydelse varierar beroende på tidpunkt och om det gäller smärta eller funktionsnedsättning som utvärderas.
ACKNOWLEDGEMENTS

Being a Ph-D student in this study, run in 6 counties, with many thousands of possible participants to consider, has been a big challenge, and I would never have imagined the amount of spare time needed to see it through. But it has also been very interesting, inspiring and rewarding, and thus become both work and “hobby” and I am so grateful for having had the opportunity to go through with it. For this to be possible however, there are many people who have contributed in various ways, both with scientific support, discussions and guidance, but also by being good and supportive family, friends and colleagues. For this I would like to express my sincere gratitude to everyone who in any way has contributed. I would also like to thank the following persons in particular.

My main supervisor, Anneli Peolsson, first of all for inviting me to be a Ph-D-student in this very interesting study, for believing in me, for your support, positive attitude and for guidance along the way. Thank you for your replies and comments, always without delay. Thank you also for introducing me to the researchers at the University of Queensland and for your personal hospitality and friendship.

My co-supervisor Åsa Dedering, for your support and for your professional and wise comments.

My co-supervisor Gun Johansson, for your kindness, positive attitude and for giving me valuable perspectives from outside the world of physiotherapy.

My fellow PhD-student in this project Gunnel Peterson, for being such a fantastic co-worker in this project and for being who you are. Thank you for all the work we have shared, for endless fruitful discussions, from anything to ultrasound and RCT-related issues but also discussions on a personal level. We have gone through so much together. Thank you for being frank, and for being a very good friend.

My previous fellow Ph-D-students (now PhDs) in the neck research group Anna Hermansen and Johanna Wibault, for all the fruitful scientific discussions we have had, all the fun we´ve had, and for being good friends. A special thanks also to Johanna for randomizing the participants.
My managers from Rehab Väst, Helén Lövborg and Ewa Rasch, first of all for supporting my wish to do this work, and for letting me take time off my clinical work. But also for continuous belief in me, and my work, and for your positive attitude all through this journey.

Statisticians Karl Wahlin and Henrik Magnusson for valuable statistical advice, and especially to Henrik for working with me creating the closest match imputation, which I never could have managed on my own.

Lars Bernfort, for valuable guidance in the world of health economics and co-authorship in paper 4.

All Ph-D students and researchers at the Division of Physiotherapy, at Linköping University, for all the valuable discussions both at seminars, and elsewhere. A special thanks to those who have taken the time to be opponents on my work at various stages throughout these years.

Shaun O’Leary for being my supervisor/contact person during my “pre-doc” visit at the Cervical Spine and Whiplash Research unit, University of Queensland, Australia. Thank you for interesting discussions and co-authorship in paper 1.

Julia Treleaven for kindly letting me share your office during my “pre-doc” visit at the Cervical Spine and Whiplash Research unit, University of Queensland, and for demonstrating the equipment you use in your research and for further cooperation regarding the dizziness data in our RCT.

Deborah Falla, for valuable contribution and co-authorship in paper 2.

The National Research School of General Practice, for generously financing much of my “pre-doc” visit at the University of Queensland, Australia, for fantastic boarding experiences in Sweden, and for everything I have learnt from you.

My family, Konny, Emma and Daniel for your support, never-ending faith in me, and for accepting that not ordinary day-time working hours, but also evenings and week-ends working with this study, often have been standard rather than an exception to the rule during these years. A special thanks to Emma also for helping out with illustrations and photos used at various stages in this study.

My parents Georg and Margareta for always believing in me and for endless support whatever I do.
My colleagues at Rehab Väst for your support and interest, and for being good colleagues and friends working with me in the clinical “real” world. A special thanks to Anne-Charlotte Engberg for your support and for keeping me updated when I have not been there.

June Daniels for, apart from being a good friend, double-checking the language in this thesis.

All other friends and relatives for showing an interest, and for being there for me. A special thanks to Merja Bergqvist, Eva Hellberg and Angela Holmberg for not letting a lot of time pass without keeping in touch, even at times when I haven’t been as good at it.

All participants contributing to any part of this study, including treating physiotherapists, WAD participants and other test-leaders.

My part in this study was supported by funding from the Swedish government through the REHSAM Foundation, the Swedish Research Council, the regional Centre for Clinical Research, the Medical Research Council of Southeast Sweden, and the County Council of Östergötland, Sweden.
REFERENCES


REFERENCES


80


Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-126782