Sick leave due to neck or shoulder pain:

Interventions, MRI assessment and prognosis

PhD dissertation

Line Thorndal Moll

Health
Aarhus University
Department of Public Health,
Section for Clinical Social Medicine and Rehabilitation
2018
Supervisors

Claus Vinther Nielsen, MD, PhD, professor
Department of Public Health, Section for Clinical Social Medicine and Rehabilitation
Aarhus University, Denmark

Merete Labriola, OT, MPH, PhD
Department of Public Health, Section for Clinical Social Medicine and Rehabilitation
Aarhus University, Denmark

Ole Kudsk Jensen, MD, PhD
Spine Centre, Diagnostic Centre, Silkeborg Regional Hospital
8600 Silkeborg, Denmark

Christina Malmose Stapelfeldt, MHSc, PhD
Department of Public Health, Section of Clinical Social Medicine and Rehabilitation
Aarhus University, Denmark

Berit Schiøttz-Christensen, MD, PhD, professor
Spine Centre of Southern Denmark, Hospital Lillebaelt Middelfart
Institute of Regional Health Research
University of Southern Denmark, Odense, Denmark
Evaluation Committee

Professor Vivi Schlünssen (chairman and moderator of the defence)
Section for Environment, Occupation and Health, Department of Public Health
Aarhus University
Bartholins Allé 2, bygning 1260
DK-8000 Aarhus C, Denmark

Professor Kirsten Fonager
Klinisk Institut, Aalborg Universitetshospital
Hobrovej 18-22
DK-9000 Aalborg, Denmark

Professor Pierre Côté
Science Building - Room 2031
North Oshawa
2000 Simcoe Street North
Oshawa, ON L1H 7K4
University of Ontario, Canada
This dissertation is based on the following studies


**Study 2.** Line Thorndal Moll, Morten Wasmod Kindt, Christina Malmose Stapelfeldt, Tue Secher Jensen. Degenerative findings on MRI of the cervical spine: an inter- and intra-rater reliability study. Accepted for publication in Chiropractic and Manual Therapies. Acceptance received on August 6th 2018.

**Study 3.** Line Thorndal Moll, Anne Mette Schmidt, Christina Malmose Stapelfeldt, Merete Labriola, Ole Kudsk Jensen, Morten Wasmod Kindt, Tue Secher Jensen, Berit Schiøttz-Christensen. Prediction of work participation within 2 years in sickness absentees with neck or shoulder pain: the contribution of demographic, patient-reported, clinical and imaging information. The final draft is approved by all authors. The manuscript will be submitted after assessment of the dissertation.

The studies are found in Appendices 1-3.
Acknowledgements

This dissertation comprises the work of my PhD studies which were performed between 2015 and 2018 while employed at the Diagnostic Centre, Silkeborg Regional Hospital, in close collaboration with and kindly hosted by DEFACTUM, Central Denmark Region.

I am greatly indebted to a number of people, and especially to my supervisors, Merete, Christina, Ole, Berit and Claus, for giving me the space and opportunity to undertake this period of scholarship. I have learned so much from all of you. My sincere thanks to Claus and Ulrich who have supported the project in both tangible and intangible ways, without which the project would not have been possible. I also owe my gratitude to Tue and Morten who invited me into the world of magnetic resonance imaging (MRI) and taught me numerous things, not only regarding MRI, but also about understanding academic issues and organisational contexts.

The study sample originates from a randomised controlled trial which was planned and conducted before I started my PhD studies. I am grateful to the study participants and to all the clinicians and researchers involved who let me use some of the data that formed the basis of this dissertation.

A special thanks to DEFACTUM for welcoming me into the research environment, and to members of the Department for Diagnostic Imaging, Silkeborg Regional Hospital, for your hospitality while performing the MRI assessments.

Thanks to my fellow ‘sprouts’ Charlotte Maria and Cecilie for contributing to a working environment with both academic discussions and laughter. A special thanks to Anne Mette for countless things, but most of all for your friendship.

I also owe my gratitude to Kirsten Fonager and Pierre Coté who agreed to evaluate my work. I am humbled and honoured for the time and effort you have spent on my work.

Last but not least, I would like to thank my parents for always having believed in me. Thanks to my family and close friends for your support and encouragement along the way. In particular, thanks to Jesper, Jakob, Emil and Ida. The greatest privilege in my life is sharing each day with you.

Line Thorndal Moll, September 2018

Funding

The work was supported by DEFACTUM, Aarhus University Denmark, Tryg Foundation, Danish Rheumatism Association, Aase and Ejnar Danielsen Foundation, and Helga and Peter Korning Foundation.
List of abbreviations

AC: Agreement by Chance
AUC: Area Under the Curve
BI: Brief Intervention
CI: Confidence Interval
CNFDS: Copenhagen Neck Functional Disability Scale
DASH: Disabilities of the Arm, Shoulder and Hand
DREAM: Danish Register for Evaluation of Marginalisation
GP: General Practitioner
IQR: Inter Quartile Range
K: Kappa
LBP: Low Back Pain
MCIC: Minimally Clinically Important Change
MDI: Multidisciplinary Intervention
NPV: Negative Predictive Value
NRS: Numeric Rating Scale
OA: Observed Agreement
PPV: Positive Predictive Value
RCT: Randomized Controlled Trial
RTW: Return To Work
TP: Tender Point
SD: Standard Deviation
s-WPS: successful Work Participation Score (i.e. WPS ≥ 75% in Weeks 30 to 104)
 u-WPS: unsuccessful Work Participation Score (i.e. WPS < 75% in Weeks 30 to 104)
VESC: Vertebral Endplate Signal Changes
WPS: Work Participation Score
Concepts used in this dissertation

**Magnetic resonance imaging** An imaging technique which does not involve ionizing radiation. It is based on the following steps: 1) a patient is placed in a magnet, 2) a radio wave is sent in and turned off, 3) the patient emits a signal which is received and 4) the received signal is used for reconstruction of the picture (1).

**Modic changes** See 'vertebral endplate signal changes'.

**Multidisciplinary** This term describes the involvement of at least two different (health care) professionals who offer distinct therapeutic modalities.

**Myelopathy** Pathologic change of the spinal cord, often caused by compression due to spinal canal stenosis.

**Neck pain** Pain located in the posterior region of the cervical spine, from the superior nuchal line to the first thoracic spinous process. It may refer to the head, the shoulder, the anterior chest wall, or the upper extremity.

**Predict / prediction** When used, these terms describe the ability to forecast the prognosis.

**Prognosis/ prognostic factor** Prognosis describes the predicted course of a given state. For instance, it may describe the predicted course of sick leave. A prognostic factor is a variable which is associated with prognosis but without making inferences about causation (2).

**Radiculopathy** A clinical assessment requiring 1) pain radiating from the neck to the upper extremity and 2) one or more positive clinical findings: diminished deep tendon reflexes, decreased muscle strength, dermatomal sensory deficits or positive foraminal compression test.

**Reliability** "The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions" (3).

**Return to work** As a concept, the meaning of 'return to work' is literal, i.e. resumption of paid work after a period of sick leave. When used as an outcome, 'return to work' will be operationalised in the relevant methods section.
**Rotator cuff disorders** See 'subacromial impingement syndrome'.

**Sick leave** Absence from work (part-time or full-time) due to work disability. People on sick leave are considered as being temporarily work disabled (as opposed to having permanent work disability).

**Signal-to-noise ratio** A measure used in MRI which compares the true signal (reflecting anatomy) to the level of background noise. In general, a low signal-to-noise ratio results in grainy images.

**Shoulder pain** Pain located around the shoulder. It may originate from the neck, from the joints (glenohumeral, acromioclavicular, or sternoclavicular), or from soft tissues adjacent to the shoulder.

**Subacromial impingement syndrome** Compression of the supraspinatus tendon between the humeral head and the acromion, coracoacromial ligament and the inferior border of the acromioclavicular joint. It is a clinical sign seen in different stages of rotator cuff disorders which constitute the most common reason for shoulder pain (4). Rotator cuff disease may include tendinitis, bursitis, tendon changes or bony degeneration (5).

**Vertebral Endplate Signal Changes** Signal intensity change in vertebral body marrow identified on MRI. How to distinguish between type 1 and type 2 is described in Appendix 4 (the evaluation manual for Study 2). Type 1 is associated with histological signs of endplate fissuring, degeneration, and granulation tissues along with the marrow being infiltrated by multiple small blood vessels (6). Type 2 is associated with histological signs of chronic repetitive trauma of the endplates and the marrow being replaced by abundant fat (6).

**Work disability** A situation where a worker is unable to stay at or return to work because of an injury or disease (7).

**Work outcomes** Research offers numerous different outcomes regarding sick leave and return to work; for instance duration of sick leave, time to four continuous weeks of self-support, work status at a given time point (at work: yes/no), work participation score (WPS) etc. As far as possible, the specific definitions are presented in reference to the relevant citations. But when necessary, the term 'work outcomes' is used as a common
term. For instance 'improved work outcomes' will be used if duration of sick leave is reduced in some studies while in others, work participation score was increased. This also applies when referring to Studies 1 and 3 (the former used time to 4 consecutive weeks of self-support and the latter WPS). In such cases, the term 'work outcomes' will also be used as a common term.

**Work participation score** A fraction yielding scores between 0% and 100%. The numerator comprises the number of working weeks while the denominator comprises the total number of follow-up weeks (8).
Figures

Figure 1. The arena in work disability prevention

Figure 2. Flowchart, Study 1

Figure 3. Flowchart, Studies 2 and 3

Figure 4. Fraction still on sick leave, Study 1 (Kaplan-Meier)

Tables

Table 1. Overview of studies, designs, sample sizes, outcome variables and statistical analyses

Table 2. Contacts with the Spine Centre, Study 1

Table 3. MRI findings and corresponding classifications, Studies 2 and 3

Table 4. Baseline characteristics, Study 1

Table 5. Inter-rater reliability estimates, Study 2

Table 6. Un-weighted kappa estimates for neural foraminal stenosis. All images vs. only images with oblique slices available, Study 2

Table 7. Intra-rater reliability estimates, Study 2

Table 8. Baseline characteristics, Study 3

Table 9. Crude odds ratios for work participation score ≥ 75% (s-WPS), Study 3

Table 10. Predictive values, specificity, sensitivity and AUC of Models 1-4, Study 3

Table 11. Cross-tabulations showing the distribution between the classification of chance and actual WPS outcomes, Study 3

Table 12. Overview of registration mechanisms in DREAM and potential types of misclassification
# Table of contents

## Summary

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>14</td>
</tr>
<tr>
<td>Methods</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>14</td>
</tr>
<tr>
<td>Conclusions</td>
<td>15</td>
</tr>
</tbody>
</table>

## Dansk resumé

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baggrund</td>
<td>16</td>
</tr>
<tr>
<td>Metode</td>
<td>16</td>
</tr>
<tr>
<td>Resultater</td>
<td>17</td>
</tr>
<tr>
<td>Konklusioner</td>
<td>17</td>
</tr>
</tbody>
</table>

## Background

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The burden of neck and shoulder pain</td>
<td>18</td>
</tr>
<tr>
<td>Work disability, sick leave and return to work</td>
<td>18</td>
</tr>
<tr>
<td>Sick leave due to neck or shoulder pain</td>
<td>19</td>
</tr>
<tr>
<td>Benefits of work</td>
<td>20</td>
</tr>
<tr>
<td>The arena in work disability prevention (stakeholders)</td>
<td>21</td>
</tr>
<tr>
<td>Interventions</td>
<td>22</td>
</tr>
<tr>
<td>Assessment of cervical spine MRI</td>
<td>23</td>
</tr>
<tr>
<td>Prognosis</td>
<td>25</td>
</tr>
</tbody>
</table>

## Aims

<table>
<thead>
<tr>
<th>Aim</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim 1</td>
<td>27</td>
</tr>
<tr>
<td>Aim 2</td>
<td>27</td>
</tr>
<tr>
<td>Aim 3</td>
<td>27</td>
</tr>
</tbody>
</table>

## Methods

<table>
<thead>
<tr>
<th>Study</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1, the RCT</td>
<td>28</td>
</tr>
<tr>
<td><strong>Design, setting and participants</strong></td>
<td>29</td>
</tr>
<tr>
<td><strong>The clinical intervention for all participants</strong></td>
<td>31</td>
</tr>
<tr>
<td><strong>The brief intervention (BI)</strong></td>
<td>32</td>
</tr>
<tr>
<td><strong>The multidisciplinary intervention (MDI)</strong></td>
<td>32</td>
</tr>
<tr>
<td><strong>Nested RCT</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>Baseline variables</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>34</td>
</tr>
<tr>
<td><strong>Analyses</strong></td>
<td>34</td>
</tr>
<tr>
<td>Study 2, the reliability study</td>
<td>36</td>
</tr>
<tr>
<td><strong>Sample and data collection</strong></td>
<td>36</td>
</tr>
<tr>
<td><strong>Readers</strong></td>
<td>37</td>
</tr>
<tr>
<td><strong>Evaluation manual, piloting and work stations</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>Variables</strong></td>
<td>38</td>
</tr>
<tr>
<td><strong>Analyses</strong></td>
<td>39</td>
</tr>
<tr>
<td>Study 3, the prognostic study</td>
<td>41</td>
</tr>
<tr>
<td><strong>Design and participants</strong></td>
<td>41</td>
</tr>
<tr>
<td><strong>Outcome variable: Work participation score (WPS)</strong></td>
<td>41</td>
</tr>
<tr>
<td><strong>Prognostic variables</strong></td>
<td>41</td>
</tr>
</tbody>
</table>
Summary

Background
Sick leave due to neck or shoulder pain is a substantial burden, both for the individual and the society. As one of the stakeholders, health care professionals perform or contribute to: interventions aimed at return to work (RTW), assessment of cervical spine magnetic resonance imaging (MRI), and assessment of work prognosis. The ideal composition of interventions has not yet been established. Few reliability studies have assessed several degenerative findings on cervical spine MRI. No studies have explored the contributions of demographics, patient-reported, clinical, and MRI information although these different types of information are often available when assessing work prognosis.

The aims of this dissertation were: 1) to evaluate the effect of a multidisciplinary intervention compared with a brief intervention in sickness absentees with neck or shoulder pain, 2) to determine the inter- and intra-rater assessment reliability of degenerative findings on cervical spine MRI, and 3) to explore the degree to which demographic, patient-reported, clinical, and MRI information contributes to the prediction of work participation.

Methods
In a secondary care setting, Study 1 was an RCT including 168 sickness absentees who met the inclusion criteria: age 18-60 years, 4-16 weeks of sick leave due to neck or shoulder pain, and fluency in Danish. Exclusion criteria were: drug addiction, primary psychiatric disorder, recent or planned spine surgery, pregnancy, and specific musculoskeletal disorders. Participants were randomly allocated to the multidisciplinary (n = 85) or brief intervention (n = 83), and follow up was 1 year. Based on registry data, the primary outcome of RTW was defined as 4 weeks of self-support and analysed by Cox proportional hazards analysis. Secondary outcomes were pain analysed by logistic regression, and disability analysed by linear regression.

Study 2 was a reliability study using 50 cervical spine MRIs chosen from those available in Study 1. An evaluation manual with classifications of findings was composed, and a chiropractor, a radiologist and a second-year resident of rheumatology independently
assessed kyphosis, disc height, disc contour, vertebral endplate signal changes, spinal canal stenosis, neural foraminal stenosis, uncovertebral osteoarthritis and zygapophyseal osteoarthritis. Inter-rater reliability was assessed pairwise and for all three readers, while intra-rater reliability was assessed for the second-year resident of rheumatology. Prevalence estimates and un-weighted kappa statistics were reported.

Study 3 was a retrospective cohort study including all the sickness absentees from Study 1 (n = 168). A successful work outcome was defined as ≥ 75% work participation score (WPS) from Weeks 30 to 104 after enrolment. Baseline variables were categorized into demographic, patient-reported, clinical, and MRI domains. Crude logistic regression analyses were used to identify prognostic variables (p<0.2), followed by multivariable analyses including the prognostic variables in a domain-wise order. For each added domain, the probability of a successful WPS was dichotomized (≥ 50% chance or not), and positive and negative predictive values, sensitivity, specificity and area under the curve (AUC) were calculated.

**Results**

In Study 1, comparison of the multidisciplinary and the brief interventions yielded no statistically significant differences with respect to RTW, pain or disability at 1-year follow up. In Study 2, mostly substantial inter-rater reliability (K ≥ 0.61) was found, and intra-rater reliability estimates were higher for all findings. In Study 3, demographic and patient-reported variables yielded a positive predictive value of 0.72, a sensitivity of 0.74, and an AUC of 0.76. Positive and negative predictive values, sensitivity, specificity and AUC did not improve notably by adding clinical and MRI information.

**Conclusions**

Firstly, the multidisciplinary intervention did not improve RTW, pain and disability compared with the brief intervention. In Study 2, the use of simple, comprehensible classifications of degenerative findings yielded mostly substantial inter-rater reliability, and substantial to almost perfect intra-rater reliability. Thirdly, clinical and MRI variables provided no additional information for the prediction of work participation compared with only demographic and patient-reported information.
Dansk resumé

Baggrund

Formålene med denne ph.d.-afhandling var 1) at sammenligne en multidisciplinær med en kort intervention blandt sygemeldte med nakke- eller skuldersmerter, 2) at vurdere inter- og intra-læser reproducerbarheden ved vurdering af MR-scanning af nakken og 3) at undersøge i hvilken grad demografisk og anamnestisk information, objektive fund og MR-fund bidrager til forudsigelse af prognosen vedr. arbejdsdeltagelse.

Metode

Studie 2 var et reproducerbarhedsstudie omfattende 50 MR-scanninger fra nakken fra Studie 1. Der blev udarbejdet en manual med klassifikationer af MR-fund, hvorefter en radiolog, en kiropraktor og en hoveduddannelseslæge i reumatologi foretog uafhængige
vurderinger af kyfose, discushøjde, discuskontur, Modic-forandringer, spinalstenose, foraminalstenose, uncovertebral artrose og facetledsartrose. Inter-læser reproducerbarhed blev vurderet parvis og for alle tre læsere, mens intra-læser reproducerbarhed blev beregnet for hoveddannelses-lægen i reumatologi. Prævalenser af positive fund samt uvægetede kappa-estimater blev rapporteret.

Studie 3 var et retrospektivt kohortestudie omfattende alle de sygemeldte fra Studie 1 (n = 168). Et succesfuldt arbejds-outcome (s-WPS) blev defineret som ≥ 75% arbejdsdeltagelse i uge 30-104 efter inklusionen i Studie 1. Baseline variable blev inddelt i fire domæner (demografisk information, anamnestisk information, objektive fund samt MR-fund). Univariat logistisk regression identificerede prognostiske variable (p < 0.2) og efterfulgtes af multivariat logistisk regression, hvor de prognostiske variable blev inkluderet domænevis. For hvert tilføjel domæne blev sandsynligheden for s-WPS dikotomiseret: ≥ 50% chance for s-WPS eller ej, og sensitivitet, specificitet, negativ og positiv prediktiv værdi (NPV og PPV) samt arealet under kurven (AUC) blev beregnet.

Resultater
I Studie 1 var der ingen statistisk signifikant forskel mellem multidisciplinær og kort intervention mht. TTA, smerter og funktionsevne ved 1 års follow-up. I Studie 2 fandtes overvejende substantiel (K ≥ 0.61) inter-læser reproducerbarhed, mens intra-læser reproducerbarheden var højere for alle MR-fund. I Studie 3 gav demografisk og anamnestisk information en PPV på 0.72, en sensitivitet på 0.74 og AUC på 0.76. Hverken sensitivitet, specificitet, NPV, PPV eller AUC blev ændret ved tilføjelse af objektive og MR-fund.

Konklusioner
Background

The burden of neck and shoulder pain

Neck and shoulder pain are common conditions. In the general population, most estimates of 1-year prevalence for neck pain range between 30% and 50% (9) while corresponding measures for shoulder pain range between 5% and 47% (10). The large variability in prevalence estimates is partly explained by differences in study populations and case definitions (9,10).

While pain as a concept is defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage (11), there are different and more specific definitions of neck pain and shoulder pain (4,12). In this dissertation, neck pain is defined as pain perceived anywhere in the posterior region of the cervical spine, from the superior nuchal line to the first thoracic spinous process (12). Neck pain may refer to the head, the shoulder, the anterior chest wall, or the upper limb (12). Shoulder pain is perceived around the shoulder and may originate from the neck, any of the three joints adjacent to the shoulder (glenohumeral, acromioclavicular and sternoclavicular joints), or adjacent soft tissues (4).

The course of neck pain is often episodic (9) and high recurrence rates (50-75% within 1-5 years) have been seen both in the general and the working populations (13,14). For shoulder pain, it has been reported that approximately 77% have symptoms after 1 month (15), 50% after 6 months (16) and 40% after 1 year (15,16). So, while both neck pain and shoulder pain will resolve spontaneously for some people, it will persist, reoccur or worsen for others. For some, the pain may cause work disability.

Work disability, sick leave and return to work

In this dissertation, the following definition of work disability is applied: "Work disability occurs when a worker is unable to stay at work or return to work because of an injury or disease" (7). It is influenced by biopsychosocial, organizational, legislative and societal factors (17) and may be divided into two categories: a) temporal work disability (sick leave) and b) permanent work disability which may be associated with (part-time) work or
dependence on benefits (18). More specifically, the focus of this dissertation is on sick leave, i.e. a situation where a worker is partly or fully absent from work due to neck or shoulder pain and not considered permanently work disabled.

Return to work will be used both as a concept and as an outcome. As a concept, the meaning is literal, i.e. resumption of full- or part-time paid work after a period of sick leave. When used as an outcome, return to work will be operationalised in the relevant methods section.

**Sick leave due to neck or shoulder pain**

Many people can and do work in spite of their pain (19). For instance, it has been reported that the annual prevalence of workers reporting activity limitation due to neck pain ranged between 11% (Norway) and 14% (Canada) (20). But for some, neck pain not only causes activity limitation; it is also associated with sick leave. In Ontario, Canada, the weighted prevalence of neck pain in lost-time claimants was 11.3% (21) while the most recent Danish record showed that workers with neck pain accounted for 16% of sick leave days (22).

Similar numbers for shoulder pain have not been estimated in Denmark, but international evidence suggests notable rates of sick leave to be associated with shoulder pain (23,24). A Dutch study found that among workers consulting their general practitioners for their initial episode of shoulder pain, 30% reported sick leave during the 6 months after the first consultation (23). In comparison, a Danish study of secondary care patients undergoing surgery for sub-acromial impingement syndrome found that 30% of the patients were on sick leave at the time of surgery and 16% received sickness benefits 1 year after surgery. This suggested a considerably higher degree of sick leave among patients with sub-acromial impingement syndrome compared with the general population, of whom 4% were on sick leave at the time of the study (24).

While sick leave is obviously a challenge to any worker, it also constitutes a burden to society as illustrated by the quantum of sick leave. It is associated with a higher risk of future disability pension (25,26), a risk that is increased with additional numbers and/or duration of absences due to sick leave, particularly for musculoskeletal disorders (27).
In Denmark, neck pain was the reason for 1.5% of disability pensions in the years 2010-2012 (22). The percentage attributed to shoulder pain was not estimated, but a large population-based study found that among people undergoing shoulder surgery, approximately 9% left the labour market within the 2-year period after surgery (28).

The cost burden on society is high: for instance, for neck pain, the average annual expenditure amounted to 917 million Danish kroner (~160 million US dollars) for healthcare and 2028 million kroner (~353 million US dollars) for productivity loss (sick leave and premature retirement in total) in years 2010-2012 (2012-equivalent kroner) (22). High expenditure is seen in all OECD countries (29), e.g. in the USA, where 9% of the total US expenditure on healthcare was attributed to neck and back pain in 2005 (30).

**Benefits of work**

While sick leave is both an individual and a societal burden, return to work is likewise expected to have positive implications at both the individual and the societal levels. In general, having a job is better for health than not having a job (7,19,31), keeping in mind that the quality of the job and the social context should be taken into account (19). A number of positive effects of work have been supported by evidence:

- From a public health perspective, work is a determinant of health equal to education, diet and exercise (7).
- Employment and socio-economic status are the main drivers of social gradients in physical and mental health (19).
- Work can be therapeutic (it can be the psychosocial vehicle to ‘recover’ from an injury/illness (32).
- Employment is (the most) important means of obtaining economic resources (19).
- Work fulfills important psychosocial needs and is pivotal to social status and identity (19).

Moreover, people with musculoskeletal disorders have reported a number of motivators for return to work: apart from the evident financial incentives, work also re-establishes a sense of normality, control and structure over their lives (33).
Summing up, sick leave constitutes a public health problem and, for most societies, there are strong incentives for reducing sick leave. It is the underlying assumption of this dissertation that - if the quality of the job and the social context are taken into account (19) - return to work is desirable for the sickness absentees and for all stakeholders affected by sick leave.

**The arena in work disability prevention (stakeholders)**

Sick leave may be initiated by a physical condition (e.g. neck or shoulder pain), but the course is often influenced by a number of stakeholders. The multitude and interplay of potentially influential factors and stakeholders were illustrated in "The arena in work disability prevention" (Figure 1 (34)).

The worker with a work disability is in the centre of the arena while each wing depicts influential systems and their representatives who may all affect the process. The bottom green wing comprises personal factors of the disabled worker, while the red wing captures the different levels of accessible health care services and stakeholders in the healthcare system. At the top of the arena, the blue wing illustrates structural layers of the workplace system and to the right, the yellow wing illustrates influential factors and stakeholders of the legislative and insurance systems. As an imaginary umbrella influencing the entire arena, the overall societal context with its culture and politics is depicted.

In brief, when assessing the study population of this dissertation, sick leave may be triggered by neck or shoulder pain (a physical factor of the personal system), but the process of sick leave depends on an interplay among all four systems as well as the overall societal context.

While acknowledging this complexity of influences and stakeholders, the overarching perspective of this dissertation is that of the healthcare system (the left wing in Figure 1). The healthcare system and its stakeholders are involved in most interventions aimed at return to work and such interventions will be investigated. Since the attending physician (the red wing representative nearest to the worker in the centre) is often a gatekeeper in sick leave management, the tasks of the physician will also be investigated. These tasks include assessment of cervical spine MRI and prediction of work prognosis.
Interventions

The focus of this dissertation is on interventions aimed at improving return to work, while pain and disability outcomes will also be covered. Such interventions are often embedded in the healthcare system or at least they involve participation of healthcare professionals from the left wing of the arena (Figure 1). Many interventions have included more than one therapeutic modality (e.g. exercises, patient education, or cognitive behavioural therapy), and the nomenclature is not unanimous in this field of research (35). In this dissertation, the term 'multidisciplinary' will be used; it describes interventions involving at least two different (healthcare) professionals who offer distinct therapeutic modalities (35)).

If there are no symptoms or signs of major structural pathology (e.g. inflammation, fracture, malignancy), it is recommended that clinicians provide reassurance and positive advice about maintenance of activity and movement (36). Such a positive, reassuring approach was also found to improve work outcomes (37-40).
In addition to reassurance provided by a health care professional, the involvement of workplaces is pivotal in interventions aimed at returning people to work. For musculoskeletal disorders, moderate quality evidence supports that workplace interventions improve time to first return to work and lasting return to work when compared with usual care (41). This was supported by strong evidence which found improved work outcomes after workplace interventions involving at least two of three components (healthcare, service coordination, work accommodation) (42). The same review (42) found moderate evidence that such interventions reduced costs related to work disability.

Together with reassurance and workplace involvement, there is strong evidence to support multidisciplinary rehabilitation programmes improving return to work in people with persistent musculoskeletal disorders (37). Much of the research has covered low back pain, and in chronic low back pain, moderate evidence supported multidisciplinary treatment being superior to physiotherapy regarding return to work (43). A positive effect on work outcomes following multidisciplinary treatment was also found for back, neck and shoulder problems (44). Regarding pain and disability outcomes, multidisciplinary treatment outperformed both physiotherapy and usual care in chronic low back pain (43).

So, based on the literature, reassurance provided by health care professionals, workplace involvement, and a multidisciplinary intervention are expected to improve return to work, pain and disability in sickness absentees with neck or shoulder pain. There is, however, still no silver bullet regarding the design of such a multidisciplinary intervention, and sickness absentees with neck or shoulder pain remain less described in the literature compared with their peers suffering from low back pain.

**Assessment of cervical spine MRI**

A key responsibility of the attending physician is to rule out the existence of major structural or other specific pathologies (36). This is primarily done on the basis of symptoms and signs (i.e. listening to the patient’s symptoms and performing a clinical examination). In the case of neck or shoulder pain, examples of such specific pathologies include fracture, inflammation, radiculopathy, myelopathy, and rotator cuff disorder.
According to guidelines, routine imaging is not recommended for neck and shoulder pain, and imaging should only be considered if it has the potential to alter management and improve outcomes of management (45,46). Advanced imaging (e.g. magnetic resonance imaging (MRI)) should be considered in cases of no improvement or deterioration of neurological status, if there is persistent severe pain and disability, and if malignancy or other major structural disease is suspected (45,46).

Despite these clinical guidelines and a relatively stable prevalence of neck and shoulder pain in Denmark, the number of cervical spine MRIs performed has increased 18% in Denmark over recent years (47) compared to only 4% increase in the prevalence of neck and shoulder pain (48,49). Increase in the number of MRIs performed has also been seen elsewhere, e.g. in Canada (50,51).

This increase may be attributed to improved access to MRI, along with both patients and physicians understanding the potential benefits of MRI. Patients believe MRI might reveal causes for their pain (52) while physicians value the non-invasiveness, the absence of ionising radiation, and the capacity to discriminate between soft tissues (53). In specific cases of neck or shoulder pain, MRI may also be requested when the clinical evaluation leaves doubt as to whether the pain originates from the neck or from the shoulder, a discrimination which is sometimes difficult based on symptoms and signs alone (4,54).

For MRI, as for all types of imaging, there are two main tasks related to its use: 1) Interpretation of the images and 2) Communication of the findings – and both of these are required (55). To achieve consistency in the communication about MRI findings, those requesting and those interpreting MRIs need to agree about both the interpretation of images and about the language applied to describe these images. The prerequisite for this is reliability which is defined as "the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions" (3).

A number of reliability studies on cervical spine MRI have found varying degrees of inter-rater reliability for different findings (56-60). Intra-rater reliability estimates were generally higher than those for inter-rater reliability (56-60). Most studies however, only assessed reliability for a few degenerative findings or compared readers with similar education and experience.
Only one reliability study was identified which assessed several degenerative findings (61). Therefore, further studies are needed. Moreover, it was necessary to determine the intra- and inter-rater reliability of cervical spine MRI assessments within this dissertation, since this was a prerequisite to assess cervical spine MRIs in the prognostic study.

**Prognosis**

Knowledge about work prognosis can guide expectations for all involved stakeholders in the process of sick leave. In neck pain, it is recommended that clinicians assess prognostic factors for delayed recovery (36). And although some physicians experience conflicting demands from different stakeholders (62), the assessment of work prognosis is a mandatory task when requested by authorities who coordinate return to work and reimburse sickness benefits. Since correct assessment of work prognosis not only guides expectations but may also have the potential to identify individuals who need special interventions or care, there is a need for the physicians responsible to be as well-informed as possible when making their assessments.

Knowledge about factors influencing work prognosis has emerged over recent decades of research. Some examples of parameters that are negatively associated with work outcomes include: increasing age (13, 64, 99, 106), psychosocial parameters (e.g. poor recovery expectations or perceived social support (63)), increasing sick leave duration (13, 14, 44, 64, 65), work-related factors (e.g. low job control and high job demands (62)) and societal/cultural factors (e.g. the generosity of the compensation benefits (62)).

But many physicians typically categorise the range of available information differently. The order of information obtained in a clinical setting is typically as follows: 1) demographics (sex and age) are already known before the clinical encounter during which 2) patient-reported information is recorded. Based on this, 3) a focused clinical examination is performed and in some cases 4) imaging (e.g. MRI) is requested. Based on the gathered information, the physician assesses work prognosis when requested. However, especially for MRI information, evidence about the prognostic value for work outcomes is scarce.

Despite the ample literature on determinants of work prognosis, no studies were identified that addressed the order of available information by exploring the contribution of
demographics, patient-reported information, clinical information and MRI information to the prediction of work prognosis.
Aims

This dissertation seeks to contribute new knowledge about interventions, MRI assessment and prognosis in sickness absentees with neck or shoulder pain seen in secondary care. In doing so, the overarching perspective was that of the healthcare system and/or the attending physician since these are key stakeholders in the process of sick leave. Applying this perspective, the overall aim was to evaluate a range of tasks performed within the healthcare system, namely interventions aimed at return to work, reliability of MRI assessments and prediction of work participation.

The specific aims were:

Aim 1
To evaluate the effect of a multidisciplinary intervention compared with a brief intervention with respect to return to work, pain and disability in workers on sick leave due to neck or shoulder pain.

Aim 2
To determine the inter- and intra-rater assessment reliability of degenerative findings (kyphosis, disc height, disc contour, vertebral endplate signal changes, spinal canal stenosis, neural foraminal stenosis, uncovertebral osteoarthritis and zygapophyseal osteoarthritis) on MRI of the cervical spine.

Aim 3
To explore the degree to which demographic, patient-reported, clinical, and MRI information contribute to the prediction of work participation in sickness absentees with neck or shoulder pain.
Methods

Table 1 presents an overview of the designs, samples, outcome variables and analyses in this dissertation. The applied methods were interwoven with the chronology of the studies. For this reason, and for the purpose of clarity, the studies will be presented according to their chronology followed by a description of the societal and legislative conditions of the studies, and their ethics approval.

Table 1. Overview of studies, designs, sample sizes, outcome variables and statistical analyses

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>N</th>
<th>Outcome variables</th>
<th>Statistical analyses</th>
</tr>
</thead>
</table>
| 1     | RCT                           | 168| **Primary outcome:** 4 weeks RTW  
**Secondary outcomes:** pain improvement ≥ 2 points (NRS) and disability (CNFDS) | Cox proportional hazards regression analysis
Survival analysis (Kaplan-Meier)  
Logistic regression analysis  
Linear regression analysis |
| 2     | Reliability study             | 50 | **Prevalence**  
Un-weighted kappa estimates | Un-weighted kappa statistics |
| 3     | Prognostic study (retrospective cohort) | 168| **Work participation score ≥ 75% in weeks 30-104 after baseline (s-WPS)**  
Sensitivity  
Specificity  
PPV  
NPV  
AUC | Crude and multivariable logistic regression analyses  
Calculations of sensitivity, specificity, PPV, NPV and AUC |

AUC: Area Under the Curve  
CNFDS: Copenhagen Neck Functional Disability Scale  
NPV: Negative Predictive Value  
NRS: Numeric rating scale  
PPV: Positive Predictive Value  
RTW: Return To Work  
RCT: Randomized Controlled Trial  
s-WPS: successful Work Participation Score  
u-WPS: unsuccessful Work Participation Score
Study 1, the RCT

**Design, setting and participants**

From May 2009 to January 2014, participants were included in a randomized controlled trial at The Spine Centre, Silkeborg Regional Hospital, Denmark. Participants were recruited from primary care settings in seven municipalities collaborating with The Spine Centre. In these municipalities, written information about the RCT was displayed in the waiting rooms of chiropractors, physiotherapists and general practitioners (GPs). The GPs were further encouraged to refer patients fulfilling the inclusion criteria: age 18-60 years, 4-16 weeks of sick leave (part-time or full-time) owing to neck or shoulder pain, and fluency in Danish. Exclusion criteria were: drug addiction, primary psychiatric disorder, recent spine surgery, plans for spine surgery, pregnancy, and inflammatory/other specific musculoskeletal disorder.

Figure 2 presents the flowchart for Study 1. Overall, 328 patients were referred and screened for eligibility. Of these, 160 could not be included, primarily because the sick leave criterion could not be fulfilled (n = 120). In total, 168 participants were included and randomly allocated to brief intervention (BI) (n = 83) or multidisciplinary intervention (MDI) (n = 85). They all completed the RCT.
METHODS

Figure 2 Flowchart Study 1
The clinical intervention for all participants

An overview of events and contacts with The Spine Centre is presented in Table 2. At their first visit to The Spine Centre, all eligible individuals were invited to participate in the study. If they agreed, written informed consent was obtained. After this, they were all examined by a rheumatologist and a physiotherapist both of whom were dedicated spine clinicians and blinded to the subsequent randomization (which took place 2 weeks later, see Table 2). The medical history was recorded by the rheumatologist, who also performed a thorough clinical examination. In a reassuring way, the rheumatologist explained the usually limited correlation between pain and degenerative spinal MRI findings.

Pharmacological treatment was adjusted if needed and laboratory tests were requested when clinically relevant. If the clinical examination revealed signs of primary shoulder pathology, ultrasonography was performed and in case of inflammation, a steroid injection was offered (n = 2, one in each group) (66). If the clinical examination revealed signs of radiculopathy, participants were informed about the likely positive prognosis and about possible referral to a spine surgeon if improvement did not occur.

Table 2. Contacts with the Spine Centre

<table>
<thead>
<tr>
<th></th>
<th>Baseline:</th>
<th>2 weeks:</th>
<th>3-4 weeks:</th>
<th>3-6 weeks:</th>
<th>12 weeks:</th>
<th>RTW plan and regular meetings with case manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical examination and advice</td>
<td>Clinical examination and advice</td>
<td>Physiotherapist follow up and randomization</td>
<td>1st meeting with case manager</td>
<td>Information about MRI findings</td>
<td>Follow up at the physiotherapist</td>
<td></td>
</tr>
<tr>
<td>MDI group</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>BI group</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

BI: Brief Intervention
MDI: Multidisciplinary Intervention
MRI: Magnetic Resonance Imaging
RTW: Return To Work
The physiotherapist performed a standardized clinical examination. Neuromuscular testing was performed and isometric neck strength was measured except in participants with radiculopathy. In this case, tests were performed using the McKenzie method which is widely used in neck pain although only supported by moderate evidence for low back pain (67). Both clinicians encouraged the maintenance of everyday activities, work and exercise.

At the follow-up visit after 3-6 weeks, the rheumatologist explained the MRI findings still using a reassuring approach. The last follow-up visit with the physiotherapist took place after 12 weeks. An effort was made to ensure coordination between all relevant stakeholders by sending copies of medical records with minimal delay (few days).

**The brief intervention (BI)**
Apart from the clinical intervention described above, those in the BI group received no further treatment. It was recommended that they return to work when possible, and if they needed further treatment or advice, they were advised to contact their GP.

**The multidisciplinary intervention (MDI)**
In addition to the clinical intervention, those in the MDI group were scheduled for meetings with individual case managers. In collaboration with these, the sickness absentees made individual RTW plans. The case managers held regular meetings with a multidisciplinary team (an occupational therapist, the rheumatologist, the physiotherapist, a social worker, a specialist in occupational medicine and, in relevant cases, a psychologist). These team members were engaged in the RTW process when relevant: 12 participants had consultations with the psychologist while the data do not describe the involvement of the other team members. The multidisciplinary team had 1-2 hours of supervision with 2-month intervals to ensure the largest possible degree of standardization in the given MDI. Meetings between the participants and their case managers were held as often as needed.

Round-table discussions were arranged at the workplace for 19 participants and in three more cases, the case manager talked to the employers via telephone. It was up to the sickness absentees to decide whether workplaces should be involved in the RTW process, something that many of them were not inclined to do. When RTW was achieved, the cases were closed. If RTW was considered unrealistic, alternative plans were made, e.g. to take
up jobs supported by the social system. If neither regular work nor jobs supported by the social system were achievable, the case was passed on to the responsible social worker in the municipality of the sickness absentee.

**Nested RCT**

Nested in the RCT was a smaller RCT comparing two different exercise programs (68). Eighty-three of the participants with nonspecific neck pain were included in the nested RCT and randomly allocated to one of two home-based exercise groups: a general physical activity group (GPA group, n = 40) and a group doing both general physical exercise and specific strength training (SST group, n = 43). No differences in pain intensity (the primary outcome) were found between the groups.

**Baseline variables**

Baseline variables were collected from comprehensive questionnaires and from the clinical examination performed by the rheumatologist prior to randomization.

The questionnaires were filled in by the study participants prior to their first meeting with the clinicians and prior to randomization. These questionnaires covered sociodemographics, education ('no education', 'brief courses', 'skilled worker', 'education < 3 years', 'education 3-4 years', 'education > 4 years', 'other'), pain intensity (11-point Numeric Rating Scale) (69), current smoking status (yes, no) and disability (Copenhagen Neck Functional Disability Scale (CNFDS) (70) for those with primarily neck complaints and Disabilities of the Arm, Shoulder and Hand (DASH) for those with primarily shoulder complaints (71)). It further covered the Örebro Musculoskeletal Pain Questionnaire (ÖMPSQ) (72). The ÖMPSQ was developed as a screening tool aimed at identifying "yellow flags" (e.g. emotional state, coping strategies, fear avoidance beliefs, and expectations of return to work). That is, it was developed as a tool for identification of patients with psychosocial risk factors for prolonged sick leave (73).

The baseline questionnaire also covered health-related quality of life (Short-Form 36), musculoskeletal comorbidity (low back pain, leg pain), sick leave duration prior to study enrollment (weeks) (13,14,44,64,65), number of previous absences due to sick leave, a question about work being the cause for pain ('yes completely', 'yes but other issues also
Contribute to pain,' no the pain is caused by other things', 'no I do not know the reason for my pain'), and part-time sick leave ('yes', 'no').

Further baseline variables were collected from the clinical examination performed by the rheumatologist. The diagnosis radiculopathy denoted a clinical assessment based on 1) radiating pain from the neck to the upper extremity and 2) one or more positive clinical findings: diminished deep tendon reflexes, decreased muscle strength, dermatomal sensory deficits or positive foraminal compression test. The diagnosis primary shoulder disorder was not based on predefined criteria but on the rheumatologist's > 30 years of clinical experience. The diagnosis was applied when a) both symptoms and signs unambiguously corresponded with shoulder pathology and b) this clinical suspicion was supported by imaging (ultrasonography or MRI of the shoulder). Non‐specific neck pain was applied when study participants reported neck pain and the clinical examination revealed no signs of specific underlying pathology (e.g. radiculopathy).

Outcomes
The primary outcome of RTW was defined as 4 consecutive weeks of financial self-support except for study participants who held jobs supported by the social system prior to their sick leave. In these cases, 4 consecutive weeks of having returned to these jobs was regarded as RTW. Data were supplied by the Danish Register for Evaluation of Marginalization (DREAM). DREAM has information about the source of income for all adult Danish citizens who have received transfer benefits at some point since 1991. Weekly recordings are made describing whether a given individual is self-supporting or receiving some kind of transfer benefits. These benefits are registered by means of 3-digit codes and ordered hierarchically which implies that if two different benefits are received in the same week, the highest-ranking code will overwrite the other. Sickness benefits have a high priority and these are only overwritten by emigration, age pensions, and death (74). If no transfer benefits are received, the individual is regarded as being self-supporting. For the purpose of this thesis, self-support is regarded as equal to having a paid job.
The secondary outcomes were:

- self-reported changes of pain intensity (Numerical Rating Scale) for all participants. This was computed by subtracting the 1-year pain score from the baseline pain score.
- self-reported measures of disability as measured by CNFDS (for non-specific neck pain and radiculopathy) and DASH (only for primary shoulder disorder).

Data on these secondary outcomes were collected by sending postal questionnaires to the participants 1 year after enrolment. In the cases of no reply, a reminder letter was sent and a final attempt to achieve the follow-up questionnaires was done by means of a telephone call. Since follow-up data on DASH were largely incomplete, this outcome was omitted (out of 21 participants with primary shoulder disorder, only nine responded of whom one was in the MDI group and eight in the BI group).

Analyses

A power calculation was carried out prior to the RCT. Assuming that there would be a 15% difference in RTW between the groups, a power \((1 – \beta)\) of 70% and a significance level of 5\% \((\alpha = 0.05)\), a study sample with 85 participants in each group was required.

Baseline variables were analyzed after exclusion of missing values using descriptive statistics. If data were not normally distributed, median values and interquartile ranges (IQR) were reported.

For the primary outcome, survival analysis was used to estimate the time to RTW. Rates of RTW in the two intervention groups were compared using Cox proportional hazards regression analysis after confirming the assumption of proportional hazards by log-minus-log plots (results not shown). Death and emigration were defined as competing risks, and the intention-to-treat principle was applied. Following the crude analyses, adjustments were made for sex, age, sick leave duration, clinical diagnoses and part-time sick leave.

For the secondary outcome of pain, the assumptions for linear regression analysis were not met. And since a minimally clinically important change (MCIC) was desired, this was defined as MCIC ≥ 2 points (yes/no) \((75,76)\) and logistic regression analysis applied.
Crude odds ratios and odds ratios adjusted for sex, age groups and baseline pain intensity were calculated.

For the secondary outcome of disability (CNFDS), the difference between intervention groups at 1-year follow up was estimated as the difference in mean CNFDS. Crude estimates and estimates adjusted for sex, age groups and baseline CNFDS score were calculated. The model was checked by diagnostic plots of the residuals (not shown).

Finally, owing to a large number of non-responders on the secondary outcomes (n = 89), an analysis of responders vs. non-responders was performed. Different tests were used depending on the type and distribution of the variable: an unpaired T test, Chi squared text, Fisher’s exact test or Wilcoxon rank-sum test.

Except for the power calculation which was performed prior to the RCT, all statistical analyses were performed by the author of this PhD dissertation who was not involved in the clinical interventions given.

**Study 2, the reliability study**

*Sample and data collection*

In agreement with recommendations for reliability studies (77), 50 MRIs were considered a reasonable sample size for this study. The predefined inclusion criteria were: availability of cervical spine MRI and a satisfactory signal-to-noise ratio as judged by an experienced radiologist. This evaluation was based on the radiologist’s 25 years of clinical MRI experience. As shown in the flowchart (Figure 3), 78 were excluded, leaving 90 cervical spine MRIs. From a list of these 90, alternate MRIs were chosen, to ensure as random a selection as possible.

The images were supplied by five different hospitals. Sagittal T1-weighted and T2-weighted sequences were available for all MRIs and axial T2-weighted sequences for 94% of the images. In addition, 82% of the MRIs had oblique T2-weighted sequences available.
Three readers assessed the MRIs. Reader A (the author of this dissertation) was a medical doctor with 9 years of postgraduate clinical experience including experience in spinal MRI assessment for clinical purposes. She had no previous formal education in MRI. Reader B was a radiologist with 25 years of clinical experience in assessing MRI. Reader C was a
METHODS

chiropractor holding 10 years of clinical and academic experience in spinal MRI. In addition, he had completed a 1-year full-time internship in spinal MRI.

Before performing the study, reader A received 2 hours of tuition from reader B covering cervical spine MRI assessment. After this, she evaluated 50 cervical spine MRIs from patients with neck pain +/- radiculopathy and completed clinical narrative reports of these. Reader B assessed the same images, corrected the reports if necessary and explained relevant issues to reader A. The images for these narrative reports were not part of the reliability study.

While blinded to demographic and clinical data as well as previous assessments, readers A, B and C independently assessed the MRIs of the full sample. This took place over a timeframe of 5-8 weeks. Reader A evaluated the images twice, and to prevent recollection of assessments, the second evaluation was performed 6 weeks after the first one.

Evaluation manual, piloting and work stations

To assist the readers, an evaluation manual (Appendix 4) with written and visual classifications of MRI findings was drafted by Reader A, followed by adjustments and approval by Readers B and C. Following this, 10 MRIs from the study sample were independently assessed by the readers. Two consensus meetings were held, with both the classifications in the evaluation manual and the practice of the readers adjusted during this piloting phase. Using radiological work stations, the MRIs were assessed using Vitrea Core (version 1.0.0.404, vital Images Inc.).

Variables

Based on the literature (59-61,78-88) and the piloting, classifications for common and degenerative MRI findings were developed. The included MRI findings are presented in Table 3, and as shown, the intention was to create simple and clinically applicable definitions (61). From these classifications, categorical (not ordinal) data were obtained.
### Table 3. MRI findings and corresponding classifications

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis</td>
<td>0</td>
<td>Normal or straightened lordosis</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Kyphosis</td>
</tr>
<tr>
<td>Disc height</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Reduced height</td>
</tr>
<tr>
<td>Disc contour</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Bulge or protrusion</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Extrusion</td>
</tr>
<tr>
<td>Spinal canal stenosis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>&gt; 50% obliteration of CSF, no cord deformity</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt; 50% obliteration of CSF with cord deformity but no signal change.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>&gt; 50% obliteration of CSF with cord deformity and signal change</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Type 1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Type 2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Type 3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Mixed type 1 and 2</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Definite osteoarthritis</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Definite osteoarthritis</td>
</tr>
<tr>
<td>Neural foraminal stenosis</td>
<td>0</td>
<td>Normal or &lt; 50% fat obliteration</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>≥ 50% fat obliteration with or without morphological changes of the nerve root</td>
</tr>
</tbody>
</table>

CSF: cerebrospinal fluid

**Analyses**

Before the reliability calculations, the prevalence of positive findings for all three readers was assessed. This tabulation of prevalence allowed for 1) assessment of sample homogeneity and 2) identification of possible systematic differences between the readers. Both these issues may affect reliability estimates (77,89). The tabulation also supplied the
estimates for observed agreement (OA) and agreement by chance (AC) that were used in the pairwise analyses of reliability. For the three-reader reliability analysis, OA was computed by calculating the number of observations with complete agreement and dividing this number with the number of assessed anatomical sites. The three-reader AC was computed by multiplying the marginal fractions (77).

For the reliability analyses, un-weighted kappa (K) statistics were used due to the categorical nature of the data. Under the condition of total independence among the readers, K is defined as

\[
K = \frac{O_{A} - AC}{1 - AC}
\]

The analyses were performed pairwise (A1B1, A1C1, B1C1 and A1A2) and for all three readers (A1B1C1). Bearing in mind the influence of prevalence on K (77,89), reliability was only computed for variables where the readers agreed on a prevalence of ≥ 10%. The nomenclature suggested by Landis & Koch (90) was used for interpretation of the K estimates:

<table>
<thead>
<tr>
<th>K value</th>
<th>Strength of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0</td>
<td>Poor</td>
</tr>
<tr>
<td>0.0-0.2</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21-0.4</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41-0.6</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61-0.8</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81-1.0</td>
<td>Almost perfect</td>
</tr>
</tbody>
</table>

Results were reported as K estimates with 95% confidence intervals along with estimates of observed agreement and agreement by chance for all findings.
Study 3, the prognostic study

Design and participants
A retrospective cohort design was used including all the sickness absentees who were enrolled in Study 1 (n = 168).

Outcome variable: Work participation score (WPS)
Like in Study 1, data on source of income were supplied by DREAM. In an attempt to capture both reintegration and maintenance of work (91), the work participation score (WPS) was chosen as the outcome. The WPS yields scores between 0% and 100%: the numerator comprises the number of working weeks and the denominator the number of follow-up weeks (8). Weeks of self-support in DREAM were considered as working weeks and for this study, weeks with state education fund grants and related benefits were also included since these transfer incomes were interpreted as being fit for work. For three individuals who were granted early retirement, the denominator comprised only the number of weeks before the first week of retirement benefits.

The follow-up period covered weeks 30-104 since the median time until the first RTW (4 consecutive weeks of self-support) was 29 weeks. A successful work outcome was defined as WPS ≥ 75% for two reasons: 1) this threshold was previously found moderately to substantially reliable compared with other common RTW measures (8) and 2) it defined an improvement in WPS compared to the mean WPS the year before enrolment (74%). This yielded a dichotomous outcome for which the following nomenclature will be used:

- Successful WPS (s-WPS) denotes WPS ≥ 75%
- Unsuccessful WPS (u-WPS) denotes WPS < 75%

Prognostic variables
Data on demographic and patient-reported information were attained from the questionnaires described under Study 1. These data covered sex, age, education, pain intensity (11-point Numeric Rating Scale) (69), number of sick-leave episodes, duration of sick leave prior to enrollment (weeks) (13,14,44,64,65), whiplash trauma (92),
psychosocial dimensions of pain (ÖMPSQ score) (72), and worker’s compensation claim (7).

Clinical variables were available from the rheumatologist’s clinical examination. Tender points were examined by applying a pressure of the thumb ~4 kg on 18 defined locations on the body. The reliability of the rheumatologist’s tender points examination has been described as excellent (93) and in accordance with common practice among clinicians, a threshold of ≥ 11 TPs was chosen (94). Radiculopathy was considered present in sickness absentee who 1) experienced radiating pain from the neck to the upper extremity and 2) had one or more of the following signs: positive foraminal compression test, decreased muscle strength, diminished deep tendon reflexes, or sensory deficits following the pattern of a dermatome.

Except for vertebral endplate signal changes (VESC), the MRI variables listed in Table 3 were assessed on disc levels C2/C3 to C7/T1. The author of this dissertation, an MD with 9 years of clinical experience, performed the assessments of the 97 available MRIs as illustrated in Figure 3. The inter-rater reliability of these assessments was substantial for all variables except zygapophyseal osteoarthritis for which moderate reliability was found. The intra-rater reliability was substantial to almost perfect (Study 2). Since the reliability of VESC was not established in Study 2, this finding was not assessed. An MRI finding was defined as present if it occurred at ≥ 1 disc level, i.e. if spinal canal stenosis was present on two disc levels, it was counted only once.

**Analyses**

For all variables expected to influence WPS prognosis, crude logistic regression analyses were performed and variables yielding $p < 0.2$ were carried forward. In the multivariable analyses, variables were included using an approach resembling the order of obtained

---

1 Bilateral locations of: suboccipital muscle insertions, anterior aspects of intertransverse spaces C5-C7, medial aspect of the supraspinatus muscle, mid upper border of trapezius muscle, second costochondral junction, 2 cm distal to the lateral epicondyle of the elbow, upper outer quadrant of the buttock, the spot posterior to the trochanteric prominence, and medial fat pad of the knee.
METHODS

information in a clinical setting. The multivariable analyses were performed stepwise which resulted in four models including the following domains of information:

- Model 1: demographics
- Model 2: demographic and patient-reported information
- Model 3: demographic, patient-reported and clinical information
- Model 4: demographic, patient-reported, clinical and MRI information

For each model, the probability of s-WPS was calculated and dichotomized, so that the sickness absentees were classified as having ≥ 50% chance of s-WPS or < 50% chance of s-WPS. By tabulation of estimated chance and actual WPS outcomes, calculation of sensitivity, specificity, positive and negative predictive values (PPV and NPV) and area under curve, AUC, was performed.

Sensitivity denotes the proportion of sickness absentees with u-WPS who were classified as having < 50% chance of s-WPS. Specificity is the proportion of sickness absentees achieving s-WPS who were correctly classified as having ≥ 50% chance. PPV is the proportion of those classified as having < 50% chance who experienced u-WPS and NPV the proportion of sickness absentees classified as having ≥ 50% chance who achieved s-WPS.

For the purpose of best possible comparability between the models, estimates were given both for the number of individuals with complete data in a given model and for the number of individuals with complete data in the adjacent model. For instance, in Model 1, estimates were given for all 168 individuals with available demographic information and for the 161 individuals with available patient-reported information in the adjacent Model 2.

In order to appraise the properties of the models, and with the particular aim of assessing whether adding domains of information improved the prediction for those with highest and lowest probabilities of s-WPS, further tabulations were made. These compared actual WPS outcomes to estimated chance, and in these tables, chance of s-WPS was divided into four categories: chance < 30%, chance 30-49%, chance 50-69%, and chance ≥ 70%.
**Nomenclature**

In Study 3, the term 'prognostic variable' was used for baseline variables which were (positively or negatively) associated with the outcome s-WPS. No inferences were made about causality (2). The terms 'predict'/'prediction' relate to the ability of the models to forecast the prognosis regarding work participation.

**Context: The Danish Labour Market and The Danish Sickness Benefit Act**

The conduct of Study 1 and interpretation of Studies 1 and 3 should be seen in the light of societal and legislative factors (17). The Danish labour market is characterized by the so-called 'flexicurity model' which implies a large degree of flexibility within the labour force: employers are allowed to fire and hire depending on the given production needs. At the same time, unemployed workers are entitled to financial security by means of unemployment benefits. Finally, an active employment policy (95) commits the authorities to offer unemployed workers activities and guidance relevant to their future employment. This responsibility is held by the municipalities.

If sick leave occurs, the Danish Sickness Benefit Act seeks to recognize needs and challenges of both the sickness absentees and the employers (96). Sickness absentees are entitled to sickness benefits if fulfilling certain criteria regarding labour market affiliation (these criteria vary among wage earners, self-employed, and unemployed) (96). This right applies irrespective of the reason for sick leave, being work-related or not. During the time of the RCT, the right for sickness benefits lasted 52 weeks. Only six participants (MDI: n = 3 and BI: n = 3) were affected by a legislative change which reduced the right for sickness benefits from 52 to 22 weeks. Below are listed the rights and obligations of the sickness absentees, the employers and the municipalities.

Where specific time periods are given, these applied during the time of Study 1 running from 2009-2014:

- A sickness absentee has no obligation to inform the employer about the specific reason for sick leave. The sickness absentee is however obliged to cooperate regarding any relevant medical treatment as well as initiatives aimed at RTW. Depending on labour market affiliation (unemployed, wage earner or self-employed) and collective
agreements, different rules ensure the sickness absentee financial compensation (in some cases sickness benefits, in some cases the usual salary depending on the collective agreement) (96).

- The employer covers the expenses (sickness benefits or usual salary) for the first 21 (2009-2012) or 30 days (2012-2014) of sick leave. After no more than 4 weeks, the employer must meet with the sickness absentee to explore opportunities of RTW, possible work modifications etc. Before 35 days have elapsed, the employer is obliged to inform the municipality about the worker on sick leave (96). This information serves two purposes:
  
  o The amount of sickness benefits is reimbursed to the employer from the municipality thereby decreasing the employer's expenses.
  o Relevant authorities at the municipality can then organize appropriate initiatives aiming to assist the person to return to work, see below.

- When receiving the information about sick leave from the employer, social workers at the municipality collect relevant information, i.e. self-report from the sickness absentee and medical information from the responsible physician. Based on this information, the sickness absentee is allocated into one of three categories (96):
  
  o Category 1 for which complete RTW is expected within 8 weeks.
  o Category 2 for which RTW is expected later than 8 weeks but without need for a multidisciplinary intervention.
  o Category 3 for which RTW is expected later than 8 weeks with a need for multidisciplinary intervention.

For categories 2 and 3, a meeting involving a municipality social worker and the sickness absentee must be arranged within the first 8 weeks of sick leave and at regular intervals thereafter.

A certain agreement (§56-agreement) in The Danish Sickness Benefit Act needs mention. This agreement applies in cases of recurrent short-term sick leave. If agreed upon, the §56-agreement entitles the employer to reimbursement of the sickness benefit amount from the first day of sick leave (not after 21/28 days as the usual practice described above). It thus exempts the employer from expenses and may exempt the worker from a sense of "guilt" for being sick (96).
**Ethics approval**

All participants signed informed consent. The RCT was registered at Current Controlled Trials, ISRCTN51739408. All studies were approved by The Danish Data Protection Agency (J. no. 1-16-02-86-16). The RCT and the prognostic study were approved by The Central Denmark Region Committees on Health Research Ethics (M-20090027) from which the reliability study was granted exemption (case no. 86, Appendix 6).
Results

Study 1

Study 1 evaluated the effect of a multidisciplinary intervention compared with a brief intervention in sickness absentee with neck or shoulder pain (97). Baseline characteristics of the study population are presented in Table 4.

<table>
<thead>
<tr>
<th>Table 4. Baseline characteristics of the study population in Study 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Age, mean (SD) in years</td>
</tr>
<tr>
<td>≤ 40 years, n (%)</td>
</tr>
<tr>
<td>Female gender, n (%)</td>
</tr>
<tr>
<td>Marital status, single n (%)</td>
</tr>
<tr>
<td>Education, n (%)</td>
</tr>
<tr>
<td>None, brief courses, other</td>
</tr>
<tr>
<td>Skilled workers, education &lt; 3 years</td>
</tr>
<tr>
<td>Education ≥ 3 years</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
</tr>
<tr>
<td>Pain intensity (0-10) last week, median (IQR)</td>
</tr>
<tr>
<td>CNFDS score, mean (SD)</td>
</tr>
<tr>
<td>DASH score, mean (SD)</td>
</tr>
<tr>
<td>OMPSQ score, n (%)</td>
</tr>
<tr>
<td>&lt;90</td>
</tr>
<tr>
<td>90-105</td>
</tr>
<tr>
<td>&gt;105</td>
</tr>
<tr>
<td>SF-36 mental health subscale, mean (SD)</td>
</tr>
<tr>
<td>Musculoskeletal comorbidity n (%)</td>
</tr>
<tr>
<td>Low back pain</td>
</tr>
<tr>
<td>Leg pain</td>
</tr>
<tr>
<td>Physician’s diagnoses, n (%)</td>
</tr>
<tr>
<td>Non-specific neck pain</td>
</tr>
<tr>
<td>Radiculopathy</td>
</tr>
<tr>
<td>Primary shoulder disorder</td>
</tr>
<tr>
<td>Sick leave duration, n (%)</td>
</tr>
<tr>
<td>≤ 12 weeks</td>
</tr>
<tr>
<td>Previous sick leaves due to neck/shoulder pain n (%)</td>
</tr>
<tr>
<td>0 previous sick leaves</td>
</tr>
<tr>
<td>1-2 previous sick leaves</td>
</tr>
<tr>
<td>3-4 previous sick leaves</td>
</tr>
<tr>
<td>&gt; 4 previous sick leaves</td>
</tr>
<tr>
<td>Is your pain caused by your work, n (%)</td>
</tr>
<tr>
<td>Answer “no”</td>
</tr>
<tr>
<td>Current part-time sick leave, n (%)</td>
</tr>
<tr>
<td>Answer “yes”</td>
</tr>
<tr>
<td>Exercise group, n (%)</td>
</tr>
<tr>
<td>General exercise</td>
</tr>
<tr>
<td>Specific exercises</td>
</tr>
<tr>
<td>Exercises for radiculopathy</td>
</tr>
</tbody>
</table>

For the primary outcome of RTW, registry data provided 100% follow up, while follow-up data on pain and disability were only available for 79 study participants. Analyses of responders versus non-responders showed that the dropout was un-skewed with respect to RTW and all baseline variables, except for the allocation to exercise groups in the nested RCT (more responders were in the general exercise group compared with non-responders). The MDI had a median duration of 4.6 months (inter-quartile range: 3.3 – 7.4) compared to 3 months in the BI group (inter-quartile range 3 – 3).

Four study participants had already achieved RTW at baseline, and therefore were excluded from the analyses. In the adjusted analyses, 14 additional study participants were excluded due to missing values for some of the variables. Hence, 164 study participants were included in the crude analyses and 150 in the adjusted analyses. Figure 4 shows the proportion of participants still on sick leave during 1-year follow up. RTW was achieved by 98 study participants (50 in the MDI group (59%) and 48 in the BI group (58%)). When using the BI as the reference, the crude hazard ratio was 0.94 (95% CI: 0.63; 1.41) and the adjusted hazard ratio was 0.84 (95% CI: 0.54; 1.31).

For the secondary outcome of pain, comparison was also made using the BI as the reference. The crude odds ratio for achieving a clinically relevant pain reduction ≥ 2 points was 1.10 (95% CI: 0.54; 2.26) and changed to 1.18 (95% CI: 0.56; 2.48) after adjustment for sex, age-groups and baseline pain intensity.

![Figure 4. Fraction still on sick leave (Kaplan-Meier)](image-url)
For the secondary outcome of disability, crude linear regression analysis showed that the mean CNFDS score after 1 year was 1.37 points (95% CI: -1.91; 4.64) higher in the MDI group compared with the BI group. This estimate changed to 1.09 points (95% CI: -2.26; 4.45) after adjustment for sex, age-groups and baseline CNFDS score.

**Study 2**

In Study 2, most of the study population was comprised of women (n = 31, 62%), and the mean age was 43.7 years. Appendix 5 presents the prevalence of positive MRI findings for all the readers. At all disc levels, the prevalence of vertebral endplate signal changes (Modic changes) was < 10% for the a priori-decided reader comparisons (A1B1, A1C1, B1C1, A1A2, and A1B1C1). Therefore, kappa statistics were not calculated for this finding. For the other MRI findings, the prevalence estimates allowed for kappa statistics including observations from one to six anatomical sites (e.g. six anatomical sites included in K statistics for neural foraminal stenosis). From Appendix 5, it seemed that the label "reduced disc height" was used more frequently by Reader C, and by Reader A in her second assessment, compared with their fellow readers. No other systematic differences were suspected based on Appendix 5.

The inter-rater reliability estimates for all reader comparisons (A1B1, A1C1, B1C1, and A1B1C1) are presented in Table 5. For the majority of the findings, the overall inter-rater reliability (A1B1C1) ranged from moderate to almost perfect (moderate to substantial for disc height, disc contour, uncovertebral and zygapophyseal osteoarthritis; moderate to almost perfect for spinal canal stenosis; and substantial to almost perfect for kyphosis and neural foraminal stenosis). For neural foraminal stenosis, exploratory analyses were performed to assess the inter-rater reliability when including only MRIs with oblique images. As can be seen from Table 6, this did not change the estimates but the confidence intervals were slightly broadened.

Table 7 shows the intra-rater reliability estimates (A1A2). These were moderate to substantial for disc height and moderate to almost perfect for spinal canal stenosis and zygapophyseal osteoarthritis. The intra-rater reliability was substantial to almost perfect for disc contour, uncovertebral osteoarthritis and neural foraminal stenosis, while for kyphosis the intra-rater reliability was almost perfect.
### Table 5. Inter-rater reliability estimates

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>n</th>
<th>Reader pair</th>
<th>Observed agreement (%)</th>
<th>Agreement by chance (%)</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kyphosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 A1B1</td>
<td></td>
<td></td>
<td>92.0</td>
<td>56.4</td>
<td>0.82 (0.75 ; 0.89)</td>
</tr>
<tr>
<td>49 A1C1</td>
<td></td>
<td></td>
<td>89.8</td>
<td>53.6</td>
<td>0.78 (0.71 ; 0.85)</td>
</tr>
<tr>
<td>49 B1C1</td>
<td></td>
<td></td>
<td>89.8</td>
<td>52.8</td>
<td>0.78 (0.71 ; 0.86)</td>
</tr>
<tr>
<td>49 A1B1C1</td>
<td></td>
<td></td>
<td>85.7</td>
<td>31.2</td>
<td>0.79 (0.73 ; 0.85)</td>
</tr>
<tr>
<td><strong>Disc height</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150 A1B1</td>
<td></td>
<td></td>
<td>92.0</td>
<td>52.8</td>
<td>0.83 (0.74 ; 0.92)</td>
</tr>
<tr>
<td>200 A1C1</td>
<td></td>
<td></td>
<td>80.0</td>
<td>52.8</td>
<td>0.58 (0.46 ; 0.69)</td>
</tr>
<tr>
<td>150 B1C1</td>
<td></td>
<td></td>
<td>77.3</td>
<td>50.0</td>
<td>0.55 (0.42 ; 0.68)</td>
</tr>
<tr>
<td>150 A1B1C1</td>
<td></td>
<td></td>
<td>74.7</td>
<td>26.4</td>
<td>0.65 (0.57 ; 0.74)</td>
</tr>
<tr>
<td><strong>Disc contour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>177 A1B1</td>
<td></td>
<td></td>
<td>76.8</td>
<td>43.4</td>
<td>0.59 (0.49 ; 0.70)</td>
</tr>
<tr>
<td>177 A1C1</td>
<td></td>
<td></td>
<td>79.7</td>
<td>43.3</td>
<td>0.64 (0.53 ; 0.74)</td>
</tr>
<tr>
<td>200 B1C1</td>
<td></td>
<td></td>
<td>80.0</td>
<td>47.6</td>
<td>0.62 (0.52 ; 0.72)</td>
</tr>
<tr>
<td>177 A1B1C1</td>
<td></td>
<td></td>
<td>68.4</td>
<td>21.7</td>
<td>0.61 (0.54 ; 0.69)</td>
</tr>
<tr>
<td><strong>Spinal canal stenosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 A1B1</td>
<td></td>
<td></td>
<td>97.0</td>
<td>76.0</td>
<td>0.88 (0.68 ; 1.00)</td>
</tr>
<tr>
<td>100 A1C1</td>
<td></td>
<td></td>
<td>91.0</td>
<td>73.5</td>
<td>0.66 (0.47 ; 0.83)</td>
</tr>
<tr>
<td>100 B1C1</td>
<td></td>
<td></td>
<td>92.0</td>
<td>74.3</td>
<td>0.69 (0.48 ; 0.86)</td>
</tr>
<tr>
<td>100 A1B1C1</td>
<td></td>
<td></td>
<td>90.0</td>
<td>63.0</td>
<td>0.74 (0.57 ; 0.86)</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too low prevalences (i.e. ≤ 10%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Uncovertebral osteoarthritis</strong>*</td>
<td>222</td>
<td>A1B1</td>
<td>90.1</td>
<td>68.0</td>
<td>0.69 (0.57 ; 0.81)</td>
</tr>
<tr>
<td>237 A1C1</td>
<td></td>
<td></td>
<td>89.0</td>
<td>68.6</td>
<td>0.65 (0.53 ; 0.77)</td>
</tr>
<tr>
<td>230 B1C1</td>
<td></td>
<td></td>
<td>87.4</td>
<td>70.9</td>
<td>0.57 (0.43 ; 0.71)</td>
</tr>
<tr>
<td>222 A1B1C1</td>
<td></td>
<td></td>
<td>83.3</td>
<td>53.0</td>
<td>0.65 (0.51 ; 0.76)</td>
</tr>
<tr>
<td><strong>Zygapophyseal osteoarthritis</strong>*</td>
<td>270</td>
<td>A1B1</td>
<td>94.8</td>
<td>74.2</td>
<td>0.80 (0.70 ; 0.90)</td>
</tr>
<tr>
<td>144 A1C1</td>
<td></td>
<td></td>
<td>87.5</td>
<td>74.9</td>
<td>0.50 (0.31 ; 0.70)</td>
</tr>
<tr>
<td>184 B1C1</td>
<td></td>
<td></td>
<td>85.9</td>
<td>78.9</td>
<td>0.33 (0.13 ; 0.53)</td>
</tr>
<tr>
<td>135 A1B1C1</td>
<td></td>
<td></td>
<td>83.0</td>
<td>61.0</td>
<td>0.56 (0.43 ; 0.70)</td>
</tr>
<tr>
<td><strong>Neural foraminal stenosis</strong>*</td>
<td>268</td>
<td>A1B1</td>
<td>90.7</td>
<td>64.1</td>
<td>0.74 (0.65 ; 0.84)</td>
</tr>
<tr>
<td>287 A1C1</td>
<td></td>
<td></td>
<td>90.2</td>
<td>64.2</td>
<td>0.73 (0.63 ; 0.82)</td>
</tr>
<tr>
<td>275 B1C1</td>
<td></td>
<td></td>
<td>87.6</td>
<td>65.8</td>
<td>0.64 (0.53 ; 0.75)</td>
</tr>
<tr>
<td>268 A1B1C1</td>
<td></td>
<td></td>
<td>84.0</td>
<td>46.0</td>
<td>0.73 (0.63 ; 0.82)</td>
</tr>
</tbody>
</table>

* n refers to the number of MRIs assessed. ** n refers to the number of disc levels assessed.
*** n refers to the number of anatomical sites assessed (by pooling right and left hand side)
Table 6. Un-weighted kappa estimates for neural foraminal stenosis.
Assessment of all images vs. assessment of only images with oblique slices available

<table>
<thead>
<tr>
<th>Reader pair</th>
<th>N</th>
<th>All images</th>
<th>N</th>
<th>Only images with oblique slices</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1B1</td>
<td>268</td>
<td>0.74 (0.65; 0.84)</td>
<td>232</td>
<td>0.74 (0.65; 0.84)</td>
</tr>
<tr>
<td>A1C1</td>
<td>287</td>
<td>0.73 (0.63; 0.82)</td>
<td>254</td>
<td>0.72 (0.62; 0.82)</td>
</tr>
<tr>
<td>B1C1</td>
<td>275</td>
<td>0.64 (0.53; 0.75)</td>
<td>233</td>
<td>0.64 (0.53; 0.76)</td>
</tr>
<tr>
<td>A1B1C1</td>
<td>268</td>
<td>0.73 (0.63; 0.82)</td>
<td>232</td>
<td>0.70 (0.62; 0.78)</td>
</tr>
</tbody>
</table>

Table 7. Intra-rater reliability estimates

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>n</th>
<th>Reader pair</th>
<th>Observed agreement (%)</th>
<th>Agreement by chance (%)</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis*</td>
<td>50</td>
<td>A1A2</td>
<td>96.0</td>
<td>59.6</td>
<td>0.90 (0.85; 0.96)</td>
</tr>
<tr>
<td>Disc height**</td>
<td>200</td>
<td>A1A2</td>
<td>84.0</td>
<td>51.5</td>
<td>0.67 (0.57; 0.77)</td>
</tr>
<tr>
<td>Disc contour**</td>
<td>174</td>
<td>A1A2</td>
<td>88.5</td>
<td>43.9</td>
<td>0.80 (0.71; 0.87)</td>
</tr>
<tr>
<td>Spinal canal stenosis**</td>
<td>50</td>
<td>A1A2</td>
<td>94.0</td>
<td>76.6</td>
<td>0.73 (0.51; 0.90)</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prevalences too low (i.e. ≤ 10%)</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis***</td>
<td>281</td>
<td>A1A2</td>
<td>90.4</td>
<td>67.0</td>
<td>0.71 (0.61; 0.81)</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis***</td>
<td>240</td>
<td>A1A2</td>
<td>90.8</td>
<td>68.8</td>
<td>0.71 (0.59; 0.82)</td>
</tr>
<tr>
<td>Neural foraminal stenosis***</td>
<td>287</td>
<td>A1A2</td>
<td>90.6</td>
<td>62.6</td>
<td>0.75 (0.66; 0.84)</td>
</tr>
</tbody>
</table>

* n refers to the number of MRIs assessed. ** n refers to the number of disc levels assessed.
*** n refers to the number of anatomical sites assessed (by pooling right and left hand side)

Study 3

Table 8 shows the baseline characteristics of both the entire study sample and the MRI sample. As can be seen, demographic, patient-reported, and clinical variables were similarly distributed in the two samples as were the intervention groups of the RCT (97). Achievement of the outcome s-WPS (a work participation score ≥ 75% in Weeks 30-104 after enrollment) was also similarly distributed; it was achieved by 47.6% in the entire study population and by 47.4% in the MRI sample.
Table 8. Baseline characteristics, Study 3

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Entire study sample</th>
<th>MRI sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>Patient-reported information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain intensity (0-10) last week, median (IQR)</td>
<td>158</td>
<td>91</td>
</tr>
<tr>
<td>Whiplash trauma, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>Sick leave duration (months), median (IQR)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>ÖMPSQ score, mean (SD)</td>
<td>161</td>
<td>93</td>
</tr>
<tr>
<td>Education ≥ 3 years, n (%)</td>
<td>155</td>
<td>88</td>
</tr>
<tr>
<td>Ongoing worker's compensation claim, n (%)</td>
<td>139</td>
<td>82</td>
</tr>
<tr>
<td>Clinical information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiculopathy, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>≥ 11 tender points, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>Intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brief intervention, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>Multidisciplinary intervention, n (%)</td>
<td>168</td>
<td>97</td>
</tr>
<tr>
<td>MRI information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyphosis, n (%)</td>
<td>97</td>
<td>23 (23.7)</td>
</tr>
<tr>
<td>Disc height reduction, n (%)</td>
<td>97</td>
<td>71 (73.2)</td>
</tr>
<tr>
<td>Bulge, protrusion or extrusion, n (%)</td>
<td>83</td>
<td>69 (83.1)</td>
</tr>
<tr>
<td>Spinal canal stenosis, n (%)</td>
<td>97</td>
<td>14 (14.4)</td>
</tr>
<tr>
<td>Neural foraminal stenosis, n (%)</td>
<td>84</td>
<td>46 (54.8)</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis, n (%)</td>
<td>83</td>
<td>39 (47.0)</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis, n (%)</td>
<td>78</td>
<td>37 (47.4)</td>
</tr>
</tbody>
</table>

Results from the crude logistic regression analyses are seen in Table 9. For those variables yielding a p < 0.2, increased odds of s-WPS were seen for male sex, radiculopathy and kyphosis, while decreased odds were seen for increased sick leave duration (months), increased ÖMPSQ score, ≥ 11 tender points, and spinal canal stenosis. The remaining variables did not meet the criterion of p < 0.2 and for this reason were not carried forward in the multivariable analyses (age, education, worker's compensation claim, whiplash, multidisciplinary intervention, disc height reduction, disc contour change, neural foraminal stenosis, zygapophyseal and uncovertebral osteoarthritis (Table 9).
The estimates of predictive values, specificity, sensitivity and AUC are presented in Table 10. Model 1 was based on sex only, i.e. equivalent to assuming that all men would have ≥ 50% chance of s-WPS and all women would have < 50% chance. The discriminatory ability of Model 1 measured by AUC was 0.58, while sensitivity was 0.76 and specificity 0.40. A sensitivity of 0.76 reflects that among all those with u-WPS, 76% were correctly identified by Model 1. A specificity of 0.40 reflects that among those achieving s-WPS, only 40% were correctly identified, leaving 60% as false positives (i.e. with unnecessary worries about those people’s work participation).

### Table 9. Crude odds ratios for work participation score ≥ 75% (s-WPS).

<table>
<thead>
<tr>
<th>Demographic information</th>
<th>N</th>
<th>Odds ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>168</td>
<td>2.13 (1.10 ; 4.13)*</td>
<td>0.03</td>
</tr>
<tr>
<td>Age (years)</td>
<td>168</td>
<td>0.99 (0.96 ; 1.03)</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Patient-reported information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain intensity (numeric rating scale)</td>
<td>161</td>
<td>0.99 (0.96 , 1.02)</td>
<td>0.38</td>
</tr>
<tr>
<td>Whiplash trauma (yes vs. no)</td>
<td>168</td>
<td>0.76 (0.31 ; 1.81)</td>
<td>0.53</td>
</tr>
<tr>
<td>Sick leave duration (months)</td>
<td>168</td>
<td>0.60 (0.46 ; 0.78)*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>ÖMPSQ score (numerical variable)</td>
<td>161</td>
<td>0.97 (0.95 ; 0.98)*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Education ≥ 3 years (yes vs. no)</td>
<td>155</td>
<td>1.22 (0.54 ; 2.73)</td>
<td>0.63</td>
</tr>
<tr>
<td>Ongoing worker’s compensation claim</td>
<td>139</td>
<td>0.65 (0.30 ; 1.40)</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>Clinical information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiculopathy (yes vs. no)</td>
<td>168</td>
<td>2.54 (1.22 ; 5.33)*</td>
<td>0.01</td>
</tr>
<tr>
<td>≥ 11 tender points (yes vs. no)</td>
<td>168</td>
<td>0.51 (0.27 ; 0.99)*</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Multidisciplinary intervention (MDI vs. BI)</strong></td>
<td>168</td>
<td>0.79 (0.43 ; 1.45)</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>MRI information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyphosis (yes vs. no)</td>
<td>97</td>
<td>2.6 (0.98 ; 6.89)*</td>
<td>0.06</td>
</tr>
<tr>
<td>Disc height reduction (yes vs. no)</td>
<td>97</td>
<td>1.07 (0.44 ; 2.64)</td>
<td>0.88</td>
</tr>
<tr>
<td>Disc contour change (bulge, protrusion or extrusion) (yes vs. no)</td>
<td>83</td>
<td>1.30 (0.40 ; 4.13)</td>
<td>0.66</td>
</tr>
<tr>
<td>Spinal canal stenosis (yes vs. no)</td>
<td>97</td>
<td>0.39 (0.11 ; 1.35)*</td>
<td>0.14</td>
</tr>
<tr>
<td>Neural foraminal stenosis (yes vs. no)</td>
<td>84</td>
<td>1.13 (0.48 ; 2.68)</td>
<td>0.78</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis (yes vs. no)</td>
<td>83</td>
<td>0.76 (0.32 ; 1.82)</td>
<td>0.54</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis (yes vs. no)</td>
<td>78</td>
<td>1.21 (0.50 ; 2.95)</td>
<td>0.68</td>
</tr>
</tbody>
</table>

s-WPS: successful work participation score
ÖMPSQ: Örebro Musculoskeletal Pain Screening Questionnaire
MRI: Magnetic Resonance Imaging
*: Variables carried forward in the multivariable analyses
<table>
<thead>
<tr>
<th>Information Included</th>
<th>N</th>
<th>Classified as ≥ 50% chance and achieved s-WPS (NPV)</th>
<th>Classified as &lt; 50% chance and achieved u-WPS (PPV)</th>
<th>s-WPS correctly classified (specificity)</th>
<th>u-WPS correctly classified (sensitivity)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1:</td>
<td>168</td>
<td>0.60 (0.46; 0.74)</td>
<td>0.58 (0.49; 0.67)</td>
<td>0.40 (0.29; 0.52)</td>
<td>0.76 (0.66; 0.85)</td>
<td>0.58</td>
</tr>
<tr>
<td>Demographic</td>
<td></td>
<td>0.61</td>
<td>0.59</td>
<td>0.41</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>Demographic *</td>
<td>161</td>
<td>0.61</td>
<td>0.59</td>
<td>0.41</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>Model 2:</td>
<td>161</td>
<td>0.70 (0.59; 0.80)</td>
<td>0.72 (0.62; 0.81)</td>
<td>0.68 (0.57; 0.79)</td>
<td>0.74 (0.63; 0.83)</td>
<td>0.76</td>
</tr>
<tr>
<td>Demographic + patient-reported **</td>
<td>161</td>
<td>0.70 (0.59; 0.80)</td>
<td>0.72 (0.62; 0.81)</td>
<td>0.68 (0.57; 0.79)</td>
<td>0.74 (0.63; 0.83)</td>
<td>0.76</td>
</tr>
<tr>
<td>Model 3:</td>
<td>161</td>
<td>0.74 (0.62; 0.84)</td>
<td>0.73 (0.63; 0.82)</td>
<td>0.67 (0.55; 0.77)</td>
<td>0.79 (0.69; 0.87)</td>
<td>0.77</td>
</tr>
<tr>
<td>Demographic + patient-reported + clinical ***</td>
<td>161</td>
<td>0.74 (0.62; 0.84)</td>
<td>0.73 (0.63; 0.82)</td>
<td>0.67 (0.55; 0.77)</td>
<td>0.79 (0.69; 0.87)</td>
<td>0.77</td>
</tr>
<tr>
<td>Model 4:</td>
<td>93</td>
<td>0.68</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Demographic + patient-reported + clinical</td>
<td>93</td>
<td>0.68</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Model 4:</td>
<td>93</td>
<td>0.67</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Demographic + patient-reported + clinical + MRI ****</td>
<td>93</td>
<td>0.67</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*: male sex OR = 2.1 (95% CI: 1.1; 4.1).
**: male sex OR = 2.0 (95% CI: 0.96; 4.2), sick leave (months) OR = 0.65 (95% CI: 0.49; 0.86), ÖMPSQ score OR = 0.97 (95% CI: 0.96; 0.99).
***: male sex OR = 1.6 (95% CI: 0.74; 3.6), sick leave (months) OR = 0.65 (95% CI: 0.49; 0.87), ÖMPSQ score OR = 0.97 (95% CI: 0.96; 0.99), radioculopathy OR = 1.74 (95% CI: 0.73; 4.1), ≥ 11 tender points OR = 0.8 (95% CI: 0.38; 1.7).
****: male sex OR = 2.8 (95% CI: 0.88; 8.9), sick leave (months) OR = 0.53 (95% CI: 0.34; 0.85), ÖMPSQ score OR = 0.98 (95% CI: 0.96; 1.0), radioculopathy OR = 1.54 (95% CI: 0.42; 5.71), ≥ 11 tender points OR = 0.8 (95% CI: 0.30; 2.2), kyphosis OR = 1.6 (95% CI: 0.45; 5.8), spinal canal stenosis OR = 0.2 (95% CI: 0.03; 1.1).

NPV: Negative Predictive Value
PPV: Positive Predictive Value
AUC: Area Under the Curve
*+: comparison with previous step is supported by similar N on both steps
When patient-reported information was added in Model 2, increased estimates were seen for negative predictive value, meaning that among those classified as having ≥ 50% chance of s-WPS, this was actually achieved by 70%. The estimate of positive predictive value also increased in Model 2, meaning that 72% had u-WPS among those who were classified as having < 50% chance. In Model 2, specificity and AUC estimates also increased compared with Model 1. In Model 3, clinical information was added which caused minor changes in negative predictive value and sensitivity. None of the values were increased by adding MRI information in Model 4.

Finally, Table 11 shows more detail regarding the discriminatory abilities of the models. Based on Model 1, nobody was classified as having < 30% or ≥ 70% chance of s-WPS. The addition of patient-reported information in Model 2 increased this discrimination, i.e. some of the sickness absentees were classified as having < 30% or ≥ 70% chance of s-WPS. This ability to discriminate between those having the lowest and highest chance was not changed by adding clinical and MRI information in Models 3 and 4 (judged by comparing the distribution of chance between the models).
### Table 11. Cross-tabulations showing the distribution between the classification of chance and actual WPS outcomes

<table>
<thead>
<tr>
<th>Information included</th>
<th>N</th>
<th>Outcome</th>
<th>N</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chance of s-WPS</td>
<td></td>
<td>s-WPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt; 30%</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30% to 49%</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50% to 69%</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 70%</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Model 1: Demographics</td>
<td>168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2: Demographic + patient-reported</td>
<td>161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3: Demographic + patient-reported + clinical</td>
<td>161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4: Demographic + patient-reported + clinical + MRI</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WPS: work participation score  
S-WPS: successful work participation score  
u-WPS: unsuccessful work participation score
Discussion

Comparison with the literature, Studies 1 and 3

The key finding of Study 1 was that RTW, pain and disability outcomes were similar in the MDI and BI groups (97). The key contribution of Study 3 was the assessment of different prognostic models: knowledge about sex, sick leave duration and ÖMPSQ score (Model 2) provided important information in the prediction of work participation, and this prediction did not improve when adding clinical and MRI information (Models 3 and 4). It is also noteworthy that work outcomes were modest in both studies. In Study 1, only 59% in the MDI group and 58% in the BI group achieved RTW within 1-year follow up (97). Follow up was longer (it reached 2 years) in Study 3, but a modest work outcome was still seen: 48% achieved s-WPS.

Remembering that sick leave is influenced by a number of systems and stakeholders (Figure 1), the findings of Studies 1 and 3 will be compared with the literature using this 'work disability arena' as a framework. More specifically, population characteristics (representing the personal system) may contribute to understanding the modest work outcomes of Studies 1 and 3 which will be discussed first, followed by a discussion of the prognostic models in Study 3, since these were also based on population characteristics.

Still remembering the 'work disability arena', properties of interventions (care provided by the health care system), work-related and legislative/societal factors will also be discussed.

Population characteristics (discussing work outcomes)

Less than 60% achieved RTW in Study 1 (97) and only 48% achieved s-WPS in Study 3, outcomes which are probably attributed to population characteristics. First and foremost, the current study population was from a secondary care setting, and such study populations generally have poorer work outcomes compared with study populations from, for example, workplaces (98). The GPs' decision regarding referral to secondary care may reflect a severity of work disability that explains the modest work outcomes.

Other population characteristics which may have affected work outcomes include pain intensity, educational level, sick leave duration prior to enrollment and psychosocial
aspects of pain. The median pain intensity was 7 (Tables 4 and 8), i.e. half the study participants scored their pain between 7 and 10 on a numeric rating scale. Increasing pain intensity was previously found to be associated with poorer work outcomes \((44,64,65,99,100)\) and hence, the high pain scores seem probable contributors to understanding the modest work outcomes.

Supported by previous findings \((64,65)\), educational level may also contribute to understanding the work outcomes. It may be harder for workers with minimal or no education to maintain appropriate work which accommodates their condition compared with workers with more education, due to a greater likelihood of manual labour. Since approximately 80% of the study population had an education of less than 3 years or no education (Tables 4 and 8), this may contribute further to understanding the work outcomes of Studies 1 and 3.

The duration of sick leave is probably also a part of the explanation, and while all study participants fulfilled the criterion of 4-16 weeks of sick leave, one quarter of them had been on sick leave for ≥ 12 weeks (Table 4). Post-hoc exploratory analyses were performed to assess RTW in Study 1 for those with shorter sick leave duration. These analyses showed that for those with sick leave ≤ 12 weeks, 65% achieved RTW. Using a threshold of ≤ 8 weeks, the corresponding measure was 72%. There were no statistically significant differences between MDI and BI in any of the exploratory analyses. Exploratory analyses were also performed for Study 3. Among those with sick leave duration ≤ 12 weeks, 55% achieved s-WPS, while this was achieved by 68% among those with sick leave duration ≤ 8 weeks, i.e. better than in the entire sample (48%). These estimates were in line with extensive literature documenting the negative influence of prolonged sick leave on work outcomes \((13,14,23,44,64,65,101,102)\). Had the inclusion criterion regarding sick leave duration been narrower (e.g. only 4-8 weeks), the MDI would probably still not have outperformed the BI, but better work outcomes would likely have been seen in Studies 1 and 3.

Likewise, the presence of psychosocial 'yellow flags' measured by the Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ) was also associated with poor work outcomes in previous studies/reviews \((72,73,103)\). In Study 1, > 90% had an ÖMPSQ
score > 90 (Table 4). This threshold has previously been reported to identify the risk of sick leave with a sensitivity > 0.8 (104,105). The tool was originally developed for identification of individuals at risk of prolonged sick leave (73), and the high ÖMPSQ scores reflect the presence of psychosocial 'yellow flags' (coping strategies, fear avoidance beliefs, expectations of RTW, and emotional state) which are likely to explain the modest work outcomes of Studies 1 and 3.

*Population characteristics (discussing the prognostic models)*

In Study 3, clinical and MRI information (Models 3 and 4) did not improve prediction of s-WPS compared to sex, sick leave duration and ÖMPSQ score alone (Model 2). For this reason, Model 2 will primarily be used for comparison with other studies. Despite a comprehensive literature search adapted from a 2017 review (106), no prognostic studies were identified which classified prognostic variables and performed statistical analyses similar to those used in Study 3. Therefore, the results could not be directly compared with other studies. Some studies did however report measures which are adequate for comparison.

In a Norwegian study (107), sickness absentees were asked to predict sick leave exceeding 26 weeks, and this self-prediction yielded a positive predictive value (PPV) of 0.78 (vs. 0.72 in Model 2). This may reflect that sickness absentees have detailed and elaborate insight into all relevant biopsychosocial aspects of their situation and furthermore, that achievement/non-achievement of given work outcomes may also be influenced by the very same sickness absentees. In contrast to the high PPV were the measures of sensitivity in the same study (107). Self-prediction yielded a sensitivity of 0.28, while sensitivity based on the assessment of dedicated medical consultants only reached 0.07. If the intent had been to identify individuals needing a certain RTW intervention, both the sickness absentees and the medical consultants in this study (107) would have identified considerably fewer individuals than Model 2 (sensitivity = 0.74). When comparing the information that was available to the medical consultants (107) with the information of Model 2, the most important difference was that sick leave duration prior to enrollment was only known in half of the cases in Fleten et al.'s study. As has already been argued, this is an important prognostic factor for work outcomes.
In another study (108), rehabilitation professionals predicted the chance of RTW, and their estimates were concordant with actual RTW in 73% of the cases, i.e. similar to the specificity of Model 2 (0.68). The rehabilitation professionals had knowledge about sex and sick leave duration like in Model 2, but their additional knowledge about the reason for sick leave, unemployment, age > 45 years and ‘gut feeling’ (108) did not elicit better specificity than in Model 2.

Without the use of any statistical models, but instead based on clinical experience, physicians were asked to describe why 195 sick leave case reports were considered particularly problematic in a qualitative study (109). Of these cases, 63% had sick leave lasting > 12 months and 29% > 3 months. Only in a minority of cases were the results of the clinical examination described (it may have been performed in more cases but was not described, hence presumably not considered important by the physicians). The results of Study 3 suggest that they correctly considered long sick leave duration as problematic and correctly offered attention to the clinical examination in only a minority of cases.

In Model 2, the variables contributing to prediction of work participation were sex, sick leave duration and ÖMPSQ score. The fact that sick leave duration and psychosocial aspects of pain measured by ÖMPSQ were predictive of s-WPS corroborated the well-established influence of these factors on work outcomes as described above (13,14,44,64,65,72,73,103,110). Regarding the impact of sex, results have been conflicting (44,64), and given the dichotomous nature of sex, the modest predictive abilities of Model 1 (based only on sex) were not surprising.

Compared with Model 2, clinical information (radiculopathy, ≥ 11 tender points) and MRI information (kyphosis, spinal canal stenosis) were added in Models 3 and 4, and neither of these models elicited notable improvements in the prediction of successful work participation. A crude association between tender points (numerical variable) and work outcomes was found in a cohort study of low back pain patients, but the association was not maintained in the multivariable model (111). The same cohort study did not find radiculopathy associated with work outcomes (111) which was corroborated by a review on subacute and chronic low back pain (106). These previous findings are in line with the limited improvement of prediction in Model 3. Regarding MRI information, only one study
was identified which examined the association with work outcomes. This study (112) suggested an association between Modic type 1 changes and unsuccessful RTW. Unfortunately, Modic type 1 changes could not be explored in Study 3 since the intra-rater and inter-rater reliability of this finding had not been established (Study 2).

Some population characteristics that are known to influence work outcomes were not included in the multivariable models of Study 3. Examples include educational level and workers’ compensation claims. Based on the odds ratios (Table 9), the influence of these variables was in line with previous findings (64,65,113), but since they did not fulfill the criterion of $p < 0.2$, they were not carried forward in the multivariable analyses. The limited sample size might explain why this criterion could not be fulfilled in the crude analyses and it cannot be refuted that a larger sample including more information in the multivariable models might have improved prediction of successful work participation.

**Components of interventions**

Understanding of the components of the interventions may contribute to knowledge as to why the MDI did not improve RTW, pain and disability compared with the BI in Study 1. Similar results were seen in a study on low back pain (114). Common to this study (114) and Study 1 was that both the BI and the MDI groups were offered the same clinical intervention by dedicated spine clinicians. The impact of this should probably not be underestimated. Secondly, the MDI had a median duration of 4.6 months which was longer than a few other interventions showing positive effect on work outcomes (115-117). If the duration of the MDI kept the participants in a ‘pending state’ while all along their probability reduced of returning to the labour market, this duration might have been counterproductive. Thirdly, and above all, a key feature in understanding the failure of the MDI seems to be the limited involvement of workplaces, a finding which governed the design of Tables A and B (Appendix 7) which offer comparisons with other RCTs. Table A presents studies with workplace involvement and Table B, studies with scarce or no workplace involvement. Beyond that, the content of the new interventions in these RCTs differed greatly: from clinical assessment and guidance resembling the BI of Study 1 (38-40), to cognitive behavioural therapy (116,118-122), self-management programmes (123,124) or dietary supplements (122). Additional information covers the surname of the
first author, publication year, pain location, sick leave duration prior to enrollment, and finally, the study results by choice of colour: Green for statistically significant improvement of work outcomes following the new intervention compared with the reference intervention, and red for studies where no such difference was found.

Despite the intervention differences, it is readily seen that the majority of RCTs with improved work outcomes actively involved the workplaces in the RTW process (115-117,125,126) which is in line with reviews assessing the effect of workplace interventions on work outcomes (41,42). In Study 1, such involvement was indeed the intent for all study participants attending the MDI, but since many of them were not inclined to involve their workplaces, it was only done in 22% of the cases.

Another interesting characteristic emerged from these RCTs, namely that with the exception of one study (116), RCTs enrolling sickness absentees with sick leave duration ≤ 3 months improved work outcomes in more cases (38-40,115,117,125,126) than did RCTs enrolling sickness absentees with a sick leave duration exceeding 3 months (114,118,120-122,127-130). Whether the intervention content or the sick leave duration explain why work outcomes were not improved in these latter studies (114,118,120-122,127-130) cannot be determined. This simply demonstrates that the influence of sick leave duration was further supported when examining these RCTs.

Work-related and societal/legislative factors
The influence of work-related factors (e.g. physically strenuous work or poor perceived support) on the process of sick leave is well-established (7,44,64) and so is the impact of societal and legislative factors (34,62,131). Since workplaces were only involved in a minority of cases, it cannot be refuted that work-related factors might partly explain the modest work outcomes of Studies 1 and 3. It may also be that societal/legislative factors have affected these work outcomes.

For instance, the limited involvement of workplaces in Study 1 was explained by the fact that many participants were not interested in such involvement. This may attribute to the Danish Act on Health Information (132), an act with the purpose of preventing health-related discrimination in workplaces. The act ensures that employers only under special
circumstances have the right to know the health conditions of their employees. In Study 1, participants may not have wanted their employers to know the specific reason for sick leave and therefore declined workplace involvement. With the available data, this cannot be unraveled, but it could represent a counterproductive interplay between legislative and workplace-related influences (the Danish Act on Health Information is meant to protect the workers, but if the employers do not know the reason for sick leave it may be difficult to provide the best help and accommodation at the workplaces).

Nor can it be elucidated or refuted whether work-related or societal factors have affected the prognosis explored in Study 3. Owing to the limited sample size and the available data, the contribution of these could not be explored.

**Comparison with the literature, Study 2**

In Study 2, primarily substantial reliability estimates were found suggesting that both the inter-rater and the intra-rater agreement was notably better than what could be expected by chance. Four key issues may explain the differences in reliability estimates when comparing the results of Study 2 with other reliability studies on MRI of the cervical spine: 1) the educational background of the MRI readers, 2) the quality of the images, 3) the use of an evaluation manual, and 4) the classification of MRI findings.

Differences in the educational background and level of experience of the MRI readers may explain why the reliability estimates were poorer in Study 2 compared with those of other studies which used readers with similar education and experience (58,59).

But similar educational background does not seem to ensure high reliability estimates. Although the MRI readers in another study had similar education and experience, the reliability estimates of this study (133) were poorer compared with Study 2. A potential explanation is a better image quality in Study 2 which has likely reduced random variation in the MRI assessments. Another possible explanation is the use of the evaluation manual which reduced both random and systematic variation in the MRI assessments.

Finally, the consequent use of the evaluation manual along with simple, comprehensible classifications of MRI findings are probable explanations why the reliability estimates of Study 2 exceeded those of other studies (60,82,134,135)
**Strengths**

The studies of this dissertation have some strengths both in terms of the chosen methods and the attempt to cover research questions that, to date, have scarcely been covered in the literature.

First, the randomised design of Study 1 was chosen because this design is the standard of excellence for studies of treatment effect, due to its prevention against both known and unknown confounders.

Second, bias was reduced in different ways:

- By way of allocation concealment, selection bias was avoided in Study 1; i.e. the allocation to intervention groups could not be affected by the clinicians, investigators or study participants.
- The study participants and the rheumatologist had no knowledge about the subsequent random allocation to intervention groups, nor did they know the outcomes of the study when providing their assessments of explanatory variables (study participants filled in the baseline questionnaires and the rheumatologist performed the clinical examination prior to randomisation). This protected against information bias in Studies 1 and 3.
- Assessment of the chosen work outcomes in Studies 1 and 3 was performed without knowledge of the intervention groups or other baseline characteristics, thus also preventing information bias.
- The use of registry data in Studies 1 and 3 ensured 100% follow up on work outcomes, thereby reducing the risk of attrition bias.
- Blinding of the three MRI readers in Study 2 served as protection against the information bias that could have threatened the estimates if the readers had possessed knowledge about the age of the study participants, their clinical data or the MRI assessments of their fellow readers.
- Blinding of the MRI reader in Study 3 protected against the information bias that could have occurred if she had known the work outcomes of the study participants when assessing the MRIs.
Third, the rigorous use of the evaluation manual in Study 2 was the prerequisite for minimising both systematic and random variation in the MRI assessments and enhancing reliability.

A fourth strength was the statistical approach in Study 3 which resembled the typical order of information sought in a clinical setting. The risk of data-driven results was reduced by the a priori decision on this statistical approach.

Fifth, exploring the impact of MRI findings on work outcomes is considered a strength since there has been a paucity of studies investigating this. Indeed, ample literature has covered the influence of psychosocial, work-related, and societal factors on the course of sick leave (7,17,62). With this in mind, the rationale for exploring a possible impact of clinical and MRI findings on WPS prognosis could be reasonably questioned. Nevertheless, the aim of Study 3 was considered relevant due to the following considerations:

- The number of cervical spine MRIs has increased disproportionately compared with the prevalence of neck and shoulder pain in Denmark (47-49). Increased use of MRI is also seen elsewhere (50,51).
- Patients believe that MRI has the capacity to reveal the truth about their situation (52).
- Since medical diagnoses are still required by the system in order to legitimise sick leave, an inappropriate focus on the medical condition rather than on psychosocial, work-related and societal issues may be observed among both sickness absentees and their health care providers.

Subsequent to the above, the explorative approach of Study 3 was considered relevant, with the potential of providing important knowledge to all stakeholders involved in the process of sick leave.

Finally, the sickness benefit legislation in Denmark underwent only a minor change in the study period. A legislative change which reduced the right for sickness benefits from 52 to 22 weeks affected only six participants (MDI: n = 3 and BI: n = 3). If more study participants had been affected, bias toward the null hypothesis could have occurred in
Studies 1 and 3. But as this was not the case, the legislative change is not suspected to have affected the results.

Limitations
When assessing the results of the current PhD dissertation, methodological limitations must be considered. Such considerations are covered in the following section: Not all of the themes are considered to have affected the estimates but are nevertheless described, because this careful consideration is necessary to assess the credibility of the results.

Sample size considerations
In Study 1, the preceding power calculation required 170 study participants (85 in each group) to detect a 15% difference in RTW given a power of 70% and a significance level of 0.05. Overall, 168 participants were enrolled (85 in MDI and 83 in BI group), but the issue of interest here is not so much the lack of two participants but rather the chosen power in the original sample size calculation. A power of 70% yields a 30% risk that a difference in RTW between intervention groups would not be detected as statistically significant. Such general considerations concerning power and the risk of type 2 errors should always be made. In the specific case of Study 1 however, and as outlined in comparison with the literature, the failure of the MDI is not suspected to represent a type 2 error.

In Study 2, a power calculation was not performed, but sample size recommendations for reliability studies were followed (n = 50) (77). If a larger sample size had been used, the confidence intervals would have been narrowed further, but substantial changes in the reliability estimates would not be expected.

Finally, the study population of Study 3 comprised all of the study participants from Study 1 (n = 168), and owing to this sample size only a limited number of prognostic variables could be included in the multivariable analyses. Had the sample size been larger, it is possible that inclusion of more variables (e.g. work-related factors) might have yielded better discriminatory abilities of the models.
Bias

• After randomisation (Table 2), blinding of the study participants with respect to treatment allocation was not performed in Study 1. Theoretically, knowledge about the treatment allocation might introduce information bias (knowledge about the expanded treatment rather than the actual MDI content itself could have caused better RTW in the MDI group – and vice versa for the BI group). Owing to the nature of the interventions in Study 1, blinding of the study participants was not possible. And since equal proportions of the two intervention groups achieved RTW, this risk of information bias is not suspected to have distorted the estimates.

• After randomisation, blinding of the clinicians in Study 1 was not possible either (Table 2). Theoretically, knowledge about the intervention group could have affected the practice and communication of the clinicians in ways that affected the study outcomes more than the content of the interventions. But as outlined above, RTW rates were similar in the two groups. For this reason, the lack of clinician blinding is not suspected to have biased the estimates of Study 1.

• In Study 2, reader A received tuition from reader B prior to performing the reliability study. This tuition could represent a potential source of bias, which was addressed by meticulous use of the evaluation manual. By following this manual, all three readers had to adjust to the research setting of Study 2 compared with their usual clinical practice. Furthermore, based on the prevalence table of Study 2, readers A and B evaluated several MRI variables with some notable differences. In conclusion, any potential bias following from the tuition seems to have been satisfactorily minimised.

Confounding

In Study 1, adjustments were made for five potential confounders. This was done out of fidelity to decisions made prior to data appraisal. However, only three of these variables (age ≤ 40 years, physician’s diagnoses and part-time sick leave) were suspected to be unevenly distributed between the intervention groups based on Table 4. In general, the attempt to address confounding in the analyses should be rendered irrelevant given the inherent protection against known and unknown confounders in the randomised design. However – especially in small study populations – there is a risk that confounding factors
will be unevenly distributed despite randomisation. In hindsight, it would have been preferable to recommend that no adjustments be made or at least only adjustments for variables that were both unevenly distributed between intervention groups and constituted potential confounders based on knowledge from the literature.

*Self-reported sick leave and register data*

While the use of DREAM has some attractive advantages regarding the outcome assessment (covered in the Strengths section), potential limitations occur in two cases:

1. **The registration at baseline**
   
   At the time of enrolment in Study 1, DREAM data were not accessible. Therefore, fulfillment of the sick leave criterion (4-16 weeks) had to rely on self-report based on which all 168 participants were sick-listed at baseline. However, when appraising DREAM data, 15 of these participants did not fulfill the criterion of 4 weeks sick leave. It could be tempting to exclude these 15 study participants from the analyses to have as ‘clean’ registry data as possible. This was done in a previous study (136) but not in Study 1 owing to the following considerations:

   - Exclusion would imply a threat to the randomised design and its inherent protection against confounding.
   - Exclusion would introduce a risk of selection bias if the distribution of exposure (allocated intervention group) and outcome (RTW) among the 15 study participants differed from the rest of the sample. This distribution was actually skewed: five of the 15 study participants did not achieve RTW, four of whom were in the BI group, while the remaining 10 study participants achieved RTW, seven of whom were in the MDI group. Hence, exclusion of the 15 study participants would have introduced attrition bias and led to underestimation of the effect of the MDI.
   - Since DREAM data were not available on the day of inclusion, the enrolment had to rely on self-report. According to clinical experience, patients in secondary care do know whether they are on sick leave or not. To further support this, the self-reported measure of sick leave during the past month has been found reliable and valid (137).
• Finally, while these 15 were not registered as having sick leave before baseline in DREAM, 11 of them were registered as having sick leave immediately after baseline. That leaves only four participants who experienced the outcome (4 consecutive weeks of self-support) immediately after baseline. By default, these four were excluded from the Cox proportional hazards analysis.

2. **Outcome assessment – the registration of short-term sick leaves**

An overview of DREAM registration mechanisms that may affect the appraisal of work outcomes are presented in Table 12 below. Due to these registration mechanisms, absences due to sick leave shorter than the employer period (21 days from 2009-2012 and 30 days from 2012-2014) are usually not registered. This implies that only sick leave exceeding 21/30 consecutive days will appear as sick leave in the registry. This introduces a risk of misclassification, namely a risk that sick leave shorter than 21/30 days does not appear in the registry. This can be illustrated by comparing two imaginary workers: X who is on sick leave for 15 continuous weeks and Y who has 15 episodes of 1-week sick leave. In DREAM, worker X will appear as having had 15 weeks of sick leave while worker Y will appear as not having been on sick leave at all. In Table 12, this potential misclassification is denoted misclassification type A.

The opposite risk of misclassification applies to workers who have a §56-agreement. For these workers, the registration procedure makes sick leave appear in the registry already from Day 1. If a worker with a §56-agreement has 15 episodes of 1-week sick leave, it will appear as 15 weeks of sick leave (i.e. no misclassification contrary to the above-mentioned example of worker Y). But a worker with a §56-agreement who has 15 individual days of sick leave will appear as having 15 weeks of sick leave in DREAM. This is explained by the hierarchy of codes (misclassification type B).
The extent to which misclassification type A has occurred can unfortunately not be unravelled. If the problem is common, the work outcomes in Studies 1 and 3 might have been overestimated. Since the potential misclassification is expected to be undifferentiated (not associated with intervention groups in Study 1 and not associated with prognostic variables in Study 3), it could potentially cause bias towards the null hypothesis.
Regarding misclassification type B, data on the distribution of §56-agreements in the study population were not available, so the impact of this potential misclassification can unfortunately not be completely elucidated. It was, however, possible to identify study participants with absences due to sick leave shorter than the employer period. Under the assumption that these weeks represented §56-agreements and hence overestimated the duration of sick leave, they were tentatively regarded as working weeks. These sensitivity analyses did not change the estimates of Studies 1 and 3, so it does not seem that misclassification type B has biased the results of these studies.

External validity

Caution is warranted when making inferences based on the results of this dissertation. The starting point of the dissertation is that of Study 1, the RCT, which, owing to the design, data collection, and analyses, is considered to have strong internal validity. The RCT has limitations, however, with respect to external validity. This is explained by two major issues:

- The enrollment of study participants depended on fulfillment of inclusion and exclusion criteria. Since approximately half of the referred individuals could not be included (Figure 2), consideration must be made as to whether those not included differ from those who were included and whether any such differences could have influenced the outcomes. For ethical reasons (lack of explicit consent), this cannot be elucidated by comparison of participants with non-participants.

- The enrollment of participants depended on the referral of eligible individuals from their GPs. Whether or not the GPs have referred only sickness absentees for whom they expected a particularly poor (or good) prognosis can unfortunately not be assessed with the available data.

Following on from these two considerations, the results of Study 1 are considered generalisable to patients with neck or shoulder pain who are seen in secondary care and who fulfill the inclusion and exclusion criteria that were applied in the RCT. Since the cohort in Study 3 comprised all RCT participants, the same considerations regarding external validity apply to Study 3.
The findings of Study 2 suggest that health care professionals can achieve agreement better than what is expected by chance, if less experienced readers receive proper training, an experienced radiologist is involved and simple classifications of findings are applied. However, the conditions of Study 2 were those of a controlled research setting, not a clinical setting. In clinical settings, there may be differences from hospital to hospital as to how closely radiologists and clinicians collaborate and thus differences may arise in their agreement on MRI. Therefore, the results of Study 2 are not necessarily generalisable to other health care professionals in other settings. To assess this, independent studies would be needed.
Conclusions

Firstly, the multidisciplinary intervention did not improve 1-year RTW, pain or disability in sickness absentees with neck or shoulder pain when compared with the brief intervention.

Secondly, inter-rater estimates showed primarily substantial reliability (K ≥ 0.61) among three health care professionals assessing degenerative findings on cervical spine MRI. For all findings, the intra-rater reliability estimates were higher.

Thirdly, the prediction of work participation was not improved when adding clinical and MRI information as compared with the prediction based on demographics and patient-reported information alone.
Implications

This dissertation contributed new knowledge regarding interventions, MRI assessment and prognosis in sickness absentees with neck or shoulder pain. Given the results of this dissertation and in the light of the scientific literature, some implications deserve consideration.

For clinical practice

First, the MDI did not improve return to work, pain or disability when compared with the BI (97). Overall, these results are in line with low back pain studies comparing the same two interventions (114,130,138) and do not support broad implementation of the MDI in its given form.

Further, clinical and MRI information provided no additional value in the prediction of work participation compared with demographic and patient-reported information alone. Discriminating between those with the lowest and highest probability of successful work participation had some limitations and hence, caution is warranted regarding use of the models in clinical practice. Moreover, Study 3 was exploratory in nature, highlighting a need for independent studies, ideally in both primary and secondary care settings. Yet, the findings are useful in the communication with both sickness absentees and other stakeholders: when assessing work participation prognosis, focus should be on demographics (sex) and patient-reported information (sick leave duration and psychosocial dimensions of pain (ÖMPSQ)) rather than on clinical and MRI information.

For research

Overall, there is still a need for research exploring the timing and intensity with which different stakeholders should be involved in RTW interventions (139). Based on the literature, there is reason to believe that a multidisciplinary intervention involving the workplace will improve work outcomes (41,42). Uniform interventions are not needed, but rather interventions that target the individual needs of the sickness absentees within the boundaries of the given legislative and societal contexts. The intervention mapping approach (previously used for developing a RTW programme in occupational LBP (140))
could offer the framework for developing such interventions for sickness absentees with neck or shoulder pain.

But all sickness absentees do not necessarily require a multidisciplinary intervention (139); some may benefit from more modest interventions and there is an obvious societal interest in achieving the best possible RTW outcomes with the least costly interventions. In Denmark, there is a need for research comparing the BI to usual treatment (i.e. what is offered by primary health care). In Norway, an intervention largely similar to our brief one caused substantially reduced sick leave rates when compared with usual care, both in the short and the long term (38-40).

The reliability estimates of Study 2 were mostly substantial (K ≥ 0.61), i.e. agreement notably exceeded the agreement that could be expected by chance. Since Study 2 was performed in a controlled research setting, confirmatory studies in clinical settings are needed. And although it is indeed important to agree on MRI interpretation, such agreement is not ‘enough’. When using MRI, one step is to ensure reliability in the assessment of images. But it is also important to further explore the associations between imaging and work outcomes. Further studies, preferably from both primary and secondary care settings, are needed to assess whether clinical and MRI information improve the prediction of work participation.
Public health perspectives

In the absence of a magic bullet regarding the content of interventions, the burden of sick leave due to neck or shoulder pain remains substantial. A key feature of addressing this public health problem is to get all players onside (34,141). That is, although they may have different values and interests, the systems and stakeholders in 'the arena in work disability prevention' need to work together towards 1) a change in maladaptive pain beliefs, 2) a behaviour change, and 3) mutual recognition in all wings of the arena, acknowledging the importance and influence of the other stakeholders.

In the general population, there is still a need for knowledge about the benign character of neck and shoulder pain, about the importance of staying active and staying at work if possible. Changing maladaptive pain beliefs may not prevent neck or shoulder pain occurring, nor does changing beliefs necessarily elicit behaviour change (30). But it would constitute a prerequisite for changing counter‐productive illness behaviour and thereby establish a prerequisite for reducing sick leave if reinforced by the other stakeholders of the arena.

In the health care system, health care professionals from different disciplines need to provide consistent information and advice about neck and shoulder pain, about causes, treatment and prognosis. As with low back pain, there is also a need for putting an end to ineffective treatments which do not reinforce healthy coping strategies and may keep patients in a counterproductive sick role (141).

Workplaces and legislative/societal structures are equally important in getting all players onside. Both should promote and support a change of pain beliefs and concurrently offer the opportunity and motivation for behaviour change, i.e. for staying at, or returning to, work under reasonable circumstances. As has previously been emphasised, laws and regulations should mirror acknowledgement of the complex interplay and influences among stakeholders (139).
References


(2) Herbert RD. Cohort studies of aetiology and prognosis: they're different. J Physiother 2014 Dec;60(4):241-244.


(92) Biering-Sorensen S, Moller A, Stoltenberg CD, Holm JW, Skov PG. The return-to-work process of individuals sick-listed because of whiplash-associated disorder: a three-year follow-up study in a


(111) Jensen OK, Stengaard-Pedersen K, Jensen C, Nielsen CV. Prediction model for unsuccessful return to work after hospital-based intervention in low back pain patients. BMC Musculoskelet Disord 2013 Apr 19;14:2474-14-140.

(112) Jensen OK, Nielsen CV, Sorensen JS, Stengaard-Pedersen K. Type 1 Modic changes was a significant risk factor for 1-year outcome in sick-listed low back pain patients: a nested cohort study using magnetic resonance imaging of the lumbar spine. Spine J 2014 Nov 1;14(11):2568-2581.


Appendices

Appendix 1 Study 1
Appendix 2 Study 2
Appendix 3 Study 3
Appendix 4 The evaluation manual used in Study 2
Appendix 5 Prevalence table of positive MRI findings in Study 2
Appendix 6 Exemption from the Ethical Committee, Study 2
Appendix 7 Overview of RCTs aimed at RTW
Study 1
Return to Work in Employees on Sick Leave due to Neck or Shoulder Pain: A Randomized Clinical Trial Comparing Multidisciplinary and Brief Intervention with One-Year Register-Based Follow-Up

Line Thorndal Moll1,2,3 · Ole Kudsk Jensen3 · Berit Schiøttz-Christensen4 · Christina Malmose Stapelfeldt1,2 · David Høyrup Christiansen5 · Claus Vinther Nielsen1,2 · Merete Labriola1,2

© The Author(s) 2017. This article is an open access publication

Abstract  Purpose The aim of this study was to evaluate the effect of a multidisciplinary intervention (MDI) compared to a brief intervention (BI) with respect to return to work (RTW), pain and disability in workers on sick leave because of neck or shoulder pain. Methods 168 study participants with sickness absence for 4–16 weeks due to neck or shoulder pain were enrolled in a hospital-based clinical study and randomized to either MDI or BI. The primary outcome was RTW obtained by a national registry on public transfer payments. Secondary outcomes were self-reported pain and disability levels. One-year follow-up RTW rates were estimated by Cox proportional hazard regression adjusted for gender, age, sick leave prior to inclusion, part-time sick leave and clinical diagnosis. Secondary outcomes were analysed using logistic and linear regression analysis for pain and disability, respectively. Results In the MDI group, 50 participants (59%) experienced four or more continuous weeks of RTW while 48 (58%) returned to work in the BI group during the 1 year of follow-up. Results showed a statistically non significant tendency towards a lower rate of RTW in the MDI group than in the BI group (adjusted HR = 0.84, 95% CI 0.54, 1.31). There were no statistically significant differences in secondary outcomes between the MDI and BI groups. Conclusion The brief and the multidisciplinary interventions performed equally with respect to both primary and secondary outcomes. The added focus on RTW in the multidisciplinary group did not improve RTW rates in this group.

Keywords  Return to work · Sick leave · Neck pain · Shoulder pain · Rehabilitation

Abbreviations
BI  Brief intervention
CNFDS  Copenhagen Neck Functional Disability Scale
DASH  Disabilities of the arm, shoulder and hand
GP  General practitioner
HR  Hazard rate
IQR  Inter quartile range
ITT  Intention to treat
MDI  Multidisciplinary intervention
MCIC  Minimally clinically important change
OR  Odds ratio
ÖMPQ  Örebro Musculoskeletal Pain Questionnaire
RCT  Randomized controlled trial
RTW  Return to work

Background
Musculoskeletal disorders are widely recognized as common causes of disability and sick leave [1–3]. Among musculoskeletal disorders, neck and shoulder pain are common,
though prevalence estimates tend to differ across studies, primarily due to differences in case definitions. In the general population, estimates of the 12-month prevalence are 2–11% for activity-limiting neck pain [3] and 5–47% for shoulder pain [4]. Among workers, 11–14% report activity limitation due to neck pain [5]. Worldwide, neck pain is the fourth most common reason for years lived with disability [1] and in Denmark, 16% of days on sick leave in 2015 were caused by neck pain [6]. Not only does sickness absence imply costs for society [7]; the potentially detrimental implications to the individual are also well described [8] as are the association between long-term sick leave and the increased risk of premature withdrawal from the labour market [9–11].

In accordance with the above, sickness absence as a focus of political concern is well established [7].

Over the past decades, the challenge of rehabilitating sickness absentees with musculoskeletal disorders has been addressed [12, 13]. Populations suffering from low back pain (LBP) are well represented in the body of literature; studies on sub-acute LBP offer moderate evidence on the positive effect of multidisciplinary rehabilitation in terms of improving disability and reducing sickness absence [14]. For chronic LBP, it is suggested based on moderate evidence that multidisciplinary rehabilitation is superior to physiotherapy with respect to return to work (RTW), pain and disability and superior to usual care with respect to pain and disability [12]. A recent review on back, neck and shoulder pain found positive RTW outcomes in studies using a multidisciplinary approach and the assignment of case managers [15]. The involvement of workplaces has also been proven beneficial [13–17]. In Denmark, the work outcomes of different studies have not been unanimous. Thus, a study from 2009 suggested positive outcomes on RTW and duration of sick leave when applying coordinated, tailored work rehabilitation in workers with musculoskeletal disorders [18]. In this study [18] however, only 19% of the participants had neck pain. More recent Danish studies evaluating work outcomes found positive effect of tailored physical activity after 3 months [19], an effect which was however not maintained at 11 months of follow-up [20]. Like in the study by Bültmann et al. [18], these studies included participants with both back, neck and shoulder pain [19, 20]. So while studies investigating pain and disability in neck and shoulder participants are common, participants with these pain locations often constitute only a minority in studies investigating work outcomes. Regarding shoulder disorders, the work outcomes of a Danish study evaluating physiotherapy exercises and occupational medical assistance are awaited [21].

In a review on the effect of different treatments for impingement syndrome1 [24] only few studies reported RTW as an outcome; neither of these fulfilled the authors’ criteria for “high quality study” and neither of these evaluated the effect of multidisciplinary interventions. Accordingly, how to rehabilitate workers on sick leave with neck and shoulder pain is a question yet to be addressed [23, 24].

Aims

The aim of this study was to evaluate the effect of a multidisciplinary intervention (MDI) compared to a brief intervention (BI) with respect to RTW, pain and disability in workers on sick leave due to neck or shoulder pain.

Methods

Design and Participants

The study was conducted as a randomized clinical trial at The Spine Centre, Silkeborg Regional Hospital, Denmark. General practitioners (GPs), physiotherapists and chiropractors in the primary sector from seven municipalities received written information about the study to display in their waiting rooms. GPs were encouraged to refer patients that fulfilled the inclusion criteria. The flow of participants is presented in Fig. 1. From May 2009 through January 2014, 328 people were screened for eligibility. Inclusion criteria were: Age 18–60 years, the primary reason for sick leave being pain in the neck, shoulders or upper thoracic region, fluency in Danish and self-reported full- or part-time sick leave for 4–12 weeks. The duration of sick leave was a pragmatic choice: patients with sick leave shorter than 4 weeks were considered to have a fairly good chance of returning to work spontaneously whereas an upper limit was chosen because longer sick leaves are associated with lower RTW chances [15]. The criterion was however changed to 4–16 weeks shortly after starting the project due to low number of referrals from GPs. Exclusion criteria were: Continuing or progressive signs of nerve root impingement implying plans for operation, known substance abuse or pregnancy, neck-, back- or shoulder-surgery within the last year, other specific or serious musculoskeletal disease and primary psychiatric disorder. Participants with comorbid psychiatric disorder considered to be in clinical remission were not excluded. 168 participants were included and completed the 1-year follow-up (Fig. 1).

Footnote 1 (continued)
coracoacromial joint. It covers a range of pathologies from inflammation of the tendon and bursa to degeneration and ultimately rupture of the tendons [22].
Randomization

An overview of the interventions is presented in Table 1. At the first visit to the Spine Centre, all participants were offered participation in the study and their written informed consent was provided. At this baseline visit, all participants were examined by a rheumatologist and a physiotherapist. Two weeks later, the first follow-up visit with the physiotherapist took place (primarily with the aim of ensuring adherence to the given exercises and making adjustments if needed) and simple randomization was carried out. A secretary made a telephone call to an externally placed computer and thereby allocated the participants to brief or multidisciplinary intervention.

Multidisciplinary Intervention Group (MDI)

In addition to the clinical examination at baseline, the participants in the MDI group had a case manager assigned who primarily had the responsibility of coordinating communication among stakeholders. Individual meetings between participants and their respective case managers were scheduled.
within 1–2 weeks after the randomization visit (Table 1). At this first meeting, they went through a standardized interview on work history, private life, pain and disability. With the aim of full or partial RTW a rehabilitation plan was made. The participant met with the case manager once or repeatedly depending on need and progress. If relevant, consultations with a psychologist were arranged (n = 12). The role as case manager was held by a social worker, a specialist of clinical social medicine or an occupational therapist. The case manager discussed relevant matters at regular team conferences not attended by the participant. Present at these team conferences were the rheumatologist, the three case managers, the physiotherapists and in relevant cases the psychologist. At the time of the study, the idea of drawing upon the expertise of the multidisciplinary team along with the access to psychologist appointments when needed was an attempt to encompass all relevant biopsychosocial considerations regarding the RTW process of the MDI group.

In 19 cases, roundtable discussions were arranged at the workplace and in three additional cases the case manager phoned the employer of the participant. The workplace involvement was optional and decided by the participants who in many cases wished to keep their health problems secret to their employers. This can be ascribed to the Danish Health Information Law [25]. In context of the Danish flexicurity model where employers have wide opportunities to “fire and hire”, the purpose of this law is to prevent discrimination of workers due to health issues. The law ensures that employers only under special circumstances are entitled to know about the health conditions of their employees. If RTW was considered impossible, an alternative plan to remain in work was made, for instance by jobs supported by the social system. To ensure a standardized multidisciplinary intervention, the entire team received 1–2 h of supervision every 2 months from a general practitioner specialized in cognitive therapy. Cases were closed when the participants returned to work and the MDI support could not proceed after this was achieved. If RTW was deemed impossible, a meeting was arranged with the municipality’s social service centre.

All Participants

Regardless of intervention group, all participants were examined by a rheumatologist and a physiotherapist at their first visit to The Spine Centre (Table 1). These two health care providers were both blinded to the subsequent random allocation to intervention groups. The rheumatologist recorded the medical history and performed a thorough clinical examination. This was followed by information about the usually limited correlation between pain and imaging of the cervical spine [26] and about aerobic exercise being beneficial for pain. Furthermore, the participants were reassured that normal daily activities, work and exercise would not be harmful. This approach was based on the findings by Indahl et al., suggesting the beneficence of reducing fear and maintaining physical activity [27]. Magnetic resonance imaging (MRI) of the cervical spine was performed except when shoulder problems were the obvious cause of pain. Participants with clinical signs of radiculopathy were informed about the good spontaneous prognosis and about the possibility of surgery in case of no improvement. If necessary, lab tests were done, and analgesic treatment was adjusted. The diagnostic accuracy of musculoskeletal ultrasound imaging has been reported moderate to high [28] for which reason participants suspected for primary shoulder disorders had ultrasound imaging of the shoulder performed. In case of ultrasonographic inflammation, a steroid injection was offered (n = 2; one in each group) [29]. The physiotherapist examined all participants in a standardised manner including neuromuscular testing and measuring isometric neck strength, except in those with radiculopathy. The latter were tested by the McKenzie method. This method is supported by moderate evidence for LBP [30] and widely used in NP though less well documented. It was none-the-less used to help participants control their pain.

At a follow-up visit approximately 3–6 weeks after enrolment (Table 1), the rheumatologist explained the MRI findings in a reassuring way and all participants had their last follow-up visit with the physiotherapist 12 weeks after their first visit.

To ensure coordination between stakeholders, copies of the medical records were sent to the participant, the GP and the municipal social services responsible for reimbursement of sick leave compensation. Except for the described follow-up visits with the rheumatologist and the physiotherapist (Table 1), those allocated to the brief intervention group were offered no further intervention. They were advised to resume work when possible. If in need for advice or additional treatment, they were recommended to consult their GP.

Nested in this randomized controlled trial (RCT) was a smaller RCT testing the effect of two different exercise programs, which has been reported previously [31]. Enrolled in the nested RCT were 83 of the participants with nonspecific neck pain who were randomly allocated to one of two home-based exercise groups. Some were allocated to a general physical activity group (GPA) (n = 40) and the remaining participants (n = 43) were allocated to a group doing both general physical exercise AND specific strength training (SST). The primary outcome of this trial was pain intensity, and no difference was found between the groups.
Context

In Denmark in the years from 2009 to 2014, when the study was conducted, a worker falling ill had the right for sick leave benefits for 52 weeks. If criteria for extending the 52 weeks were not fulfilled, only some citizens could receive other social transfer benefits from their municipality [32] since the right to other transfer benefits depended—among other things—on the spouse’s income.

Variables and Outcomes

Baseline data were collected from a questionnaire completed by the participants prior to the clinical examination. This questionnaire covered socio-demographic factors, health issues, and work-related factors. Pain intensity was measured on an 11-point numeric ranking scale from 0 (no pain) to 10 (worst imaginable pain) [33], and psychosocial dimensions of pain were measured by the Örebro Musculoskeletal Pain Questionnaire (ÖMPQ) [34, 35]. For participants with primary shoulder disorder, disability was measured by disabilities of the arm, shoulder and hand (DASH) [36] and for the rest of the study population by the Copenhagen Neck Functional Disability Scale (CNFDS) [37]. Mental health was measured by the SF-36 mental health subscale [38]. The duration of sick leave was dichotomized at a cutoff value of 12 weeks [14, 39].

The primary outcome RTW was defined as the first period of four consecutive weeks of self-support for individuals who were self-supporting before their sick leave. For those individuals who held jobs supported by the social system prior to their sick leave, four consecutive weeks of return to this job was defined as RTW. The choice of 4 weeks was explained by the wish to ensure comparability with the previously conducted LBP study [40] at The Spine Centre. RTW and sick leave compensation data were attainable from the Danish Register for Evaluation of Marginalisation (DREAM)—a national registry on public social and health-related benefits registered on a weekly basis and administered by The Danish Ministry of Employment. Since July 1991, all Danish citizens having received any type of social or health-related benefits are registered in DREAM. The source of income is registered by means of a 3-digit code and ordered hierarchically [41].

One year after inclusion, postal questionnaires were sent to the participants. These questionnaires provided data on the secondary outcomes: changes in pain level (numeric ranking scale) [33] and disability level as measured by the CNFDS [37] (participants with primary shoulder disorder excluded from the analysis). Changes in pain levels were calculated by subtracting 1-year follow-up pain levels from baseline levels. Due to a large proportion of non-responders leaving only nine participants with primary shoulder disorder with follow-up disability measures (DASH) (MDI n = 1, BI n = 8), this outcome measure was omitted.

Analyses

Prior to the study, a power calculation was carried out based on the assumption that there would be a 15% difference in RTW between the groups. Given a power (1-β) of 70%, a sample size of 85 in each group was required (two-sided α = 0.05).

The distribution of baseline characteristics was presented after excluding missing values. For those variables not fulfilling the assumption of normality, median values and inter quartile ranges (IQR) were reported.

The time to RTW during 1 year of follow-up was estimated using survival analysis (Kaplan–Meier). RTW rates in the two groups were compared using Cox proportional hazard regression. Competing risks were defined as death and emigration. The assumption of proportional hazards was assessed and confirmed using log-minus-log plots (not shown). Crude and adjusted hazard ratios (HR) were calculated according to the intention to treat (ITT) principle with adjustment for known prognostic variables for RTW: sex, age (≤40/>40 years) and duration of sick leave (≤/>12 weeks) [39, 42] as well as part-time sick leave (yes/no) and clinical diagnoses (non-specific neck pain, radiculopathy, primary shoulder disorder).

For the secondary outcome pain; two-way scatter plots (not shown) could not justify the assumption of linearity between follow-up and baseline scores. Furthermore, a minimally clinically important change (MCIC) defined as ≥2 points (yes/no) [43, 44] was considered relevant and hence, data on pain intensity changes were dichotomized according to this. Logistic regression analysis estimating crude odds ratio (OR) and adjusted OR (gender, age groups (≤/>40 years) and baseline pain intensity) was performed. To our knowledge there is no consensus on a cutoff value for a MCIC for the secondary outcome disability as measured by CNFDS. And as the model for linear regression adjusting for gender, age groups and baseline CNFDS values was checked and accepted by diagnostic plots of the residuals, this outcome measure was calculated by linear regression analysis. Positive values of β reflect increased disability levels. Due to the risk of over-fitted models in the secondary outcome analyses, the number of potential confounders was reduced to three variables compared to five in the analyses of time to RTW.

For those individuals lost to follow-up on the secondary outcomes (n = 89), a non-response analysis of responders versus non-responders was performed comparing the allocation to intervention groups, achievement of the primary outcome and all baseline characteristics (data not shown). These analyses were performed using an unpaired T test,
Fisher’s exact test, Chi squared test ($\chi^2$) or the Wilcoxon rank-sum test, depending on type and distribution of the variable. The statistical software package STATA 13.1 was used for analysis and p values $<$0.05 were regarded as statistically significant. Statistical analyses were performed by researchers outside the hospital and independently from those who gave the interventions.

Ethical Approval

All participants signed informed consent. The study is registered at Current Controlled Trials, ISRCTN51739408. It was approved by The Danish Data Protection Agency (J. No. 2007-58-0010) and by the regional ethical committee (M-20090027).

Results

Study Population

After inclusion of 168 participants, the study was closed in January 2014 primarily due to changes in the data management unit making it impossible to continue the same method of randomization, secondarily due to recruitment difficulties.

Table 2 shows baseline characteristics of the study participants. The access to register data on the primary outcome allowed for 100% follow-up, whereas a considerable dropout rate (n = 89) was seen on the secondary outcomes gathered by questionnaires. A non-response analysis revealed no differences between responders and non-responders regarding allocation to intervention group, achievement of the primary outcome or any of the baseline variables except for allocation to exercise groups (among responders, more participants were in the general exercise group compared to non-responders).

Primary Outcome: RTW

For the primary outcome RTW the total number of events was 98 and the total follow-up time was 5492 weeks. At baseline, four individuals had already experienced the event RTW and were therefore excluded from the analysis as were an additional number of 14 individuals due to missing values in one or more of the variables that we adjusted for (Fig. 1). Thus, 164 and 150 individuals were included in the crude and adjusted analyses, respectively. None of the participants were excluded due to competing risks (death and emigration).

The proportion of participants in the two groups still on sick leave is illustrated in Fig. 2. In the MDI group, 50 participants (59%) returned to work during the 1-year follow-up while 48 participants (58%) in the BI group experienced the event. The crude HR was 0.94 (95% CI 0.63; 1.41) and the adjusted HR was 0.84 (95% CI 0.54; 1.31). The median time to RTW was 44 weeks (IQR 18–52) in the MDI group and 32 weeks (IQR 12–52) in the BI group (p = 0.83). The median duration of the MDI intervention was 4.6 months (IQR 3.3–7.4) and 3 months (IQR 3–3) in the BI group.

Secondary Outcomes: Pain Intensity and Disability

The median pain score reduction was 2 units in both groups (MDI group IQR 0; 3, BI group IQR 0; 5). However, when comparing the MDI to the BI group, the crude OR for a clinically important pain reduction ≥2 points was 1.10 (95% CI 0.54; 2.26). Adjustment for gender, age-groups and baseline pain intensity yielded an OR of 1.18 (95% CI 0.56; 2.48). For disability, linear regression analysis yielded crude estimates of a non-significant CNFDS beta coefficient of 1.37 (95% CI −1.91; 4.64) points higher in the MDI group compared to the BI group at 1-year follow-up. After adjustment for gender, age-groups and baseline level of disability the coefficient changed, however still non-significantly, to 1.09 points (95% CI −2.26; 4.45) at follow-up.

Discussion

Two main findings from this study warrant exploration. One is the lack of difference between a multidisciplinary intervention compared to a brief intervention with respect to RTW, pain and disability in sick-listed workers with neck or shoulder pain. The other is the discouraging fact that less than 60% of the study population returned to work during the first year.

As for the lack of difference between the MDI and the BI; the study conducted by Bültmann et al. [18] reported a significant improvement in RTW status at 1-year follow-up in a Danish study on sick-listed workers with musculoskeletal disorders. Some notable differences in interventions and study populations may explain why we did not find similar results. The involvement of workplaces was a key element as 45% of participants in the intervention group had round-table discussions arranged at the workplace in Bültmann’s study. Also, a maximum duration of the intervention equivalent to 3 months was settled on. The mean duration of sick leave prior to the intervention was approximately 6 weeks [18]. In the present study, only 19 (22%) in the MDI group had roundtable discussions arranged, sick leave was longer and the median duration of the MDI was 4.6 months (IQR 3.3–7.4).

Another possible explanation for the lack of difference between the MDI and the BI groups could be the similarities of the clinical services provided by the rheumatologist and the physiotherapist. The approach to the
participants in both groups was based on a non-injury model as inspired by Indahl et al. [27] and Hagen et al. [45]. Both Myhre et al. [46] and Brendbekken et al. [47] had the same similarities between control and intervention groups. They did not find differences in RTW outcomes either. The reassurance provided by thorough examinations and explanations from two clinicians dedicated to spine disorders should probably not be underestimated—a point which has also previously been stated [40, 45].

Less than 60% of the participants returned to work during follow-up which is inferior to the results from similar studies describing RTW for more than 70% of their participants [18, 40, 46], and the modest RTW results warrant exploration.
In the randomized trials by Jensen et al. [40] and Myhre et al. [46], a multidisciplinary intervention much similar to the one used in the present study was offered; both reported successful RTW for approximately 70% of their participants. Differences in pain location might be an explanation, as only LBP patients were included in the former [40] whereas in the latter [46], both neck and back pain patients were included; however, the distribution of pain locations is not presented. In the above mentioned study by Bültmann [18], only 12% of the study population had neck pain. Recent studies by Andersen et al. [19, 20] found promising RTW results of tailored physical activity at 3 month follow-up but these were not maintained at 11 month follow-up; neither the tailored physical activity program nor the pain self management program improved RTW compared to the reference group. The outcome measure in these studies was RTW status (yes/no) and although different from the present four consecutive weeks of RTW [48], the proportion of participants returning to work was closer to our results than in the studies by Jensen et al. [40] and Myhre et al. [46]. A possible explanation could be a larger proportion of the study population suffering from neck and upper extremity pain. However, this information was not provided by Andersen et al.

While involvement of workplaces should be a key element in the process of RTW [15–17, 39, 42], our RTW results were notably poorer compared to the previously published LBP study by Jensen et al., although the rehabilitation programs were very similar [40]. In contrast to the previously mentioned studies [18, 40, 46], the present study included only participants with neck and shoulder pain. This may lead to considerations of the possibility of a poorer RTW prognosis for people with neck and shoulder pain in general compared to people with LBP.

Apart from the pain location, the present study population also had baseline characteristics that might have influenced the process of returning to work. At inclusion, the participants were troubled by severe pain intensity and considerable psychosocial impact of their pain (ÖMPQ) (Table 2). Both high pain intensity scores and ÖMPQ scores >90 have been shown to predict future sick leave [15, 34, 49] and thus may have affected RTW outcomes. At baseline, almost half of the study population had musculoskeletal comorbidity and approximately one-third had ≥3 previous sick leaves. Both factors are known to have negative prognostic value with respect to RTW [15, 42].

In studies with RTW outcomes similar to ours, explanations may also in part be found in baseline characteristics. Thus, in Andersen et al.’s studies [19, 20] where approximately 60% returned to work, more than half of the study population had previous sick leave episodes. In the study by Brendbekken et al. [47], the mean duration of sick leave prior to inclusion was 147 days. Both number of previous sick leaves and current sick leave duration are negative prognostic factors for RTW [15].

The study had several strengths. One was the randomized design which ensured comparability between the two groups with the exception of a larger proportion of part-time sick-listed participants in the MDI group compared to the BI group. However, this variable was adjusted for. Second, we had 100% follow up on the primary outcome thus eliminating the risk of attrition bias. A third strength of the study was the ITT analysis. The fact that baseline clinical examinations had 100% follow up on the primary outcome thus eliminated the risk of attrition bias. A third strength of the study was the ITT analysis. The fact that baseline clinical examinations were carried out blindedly before randomization was considered a further strength.

The study also had some limitations. First, given the nature of the interventions, it was not possible to perform all interventions in a blinded manner. A second potential weakness was the recruitment of participants. The GPs received written information about the study with encouragement to refer patients on sick leave due to neck and shoulder pain. They may have referred only high-risk patients because they would consider it more cost-effective to treat low-risk patients in primary care. Whether GPs have had such considerations is unknown. Although the referral pattern was similar to the LBP study [40] this aspect needs to be taken into account when considering generalizability of the study.

Third, participants with sickness absence lasting 4–16 weeks were included although longer sickness spells constitute an independent risk factor of not returning to work [15, 39]. An exploratory analysis to test if a more rigid inclusion criterion on sick leave (4–8 weeks) would have yielded different results was performed; this was not the case (data not shown). Fourth, the number of non-responders on the secondary outcomes was substantial (n = 89) introducing a potential risk of selection bias in the assessment of secondary outcomes. Non-response analysis (data not shown) did not show any statistically significant differences between responders and non-responders with

![Fig. 2 Reduction in proportion of participants on sick leave during follow-up (Kaplan–Meier)](image-url)
respect to intervention groups, RTW or any of the baseline variables. Only the allocation to exercise groups differed between responders and non-responders. This was a difference not suspected to have biased the estimates of the secondary outcomes. Nor do we, to the best of our knowledge, consider the nested RCT [31] to threaten the estimation of the results in the present study. We base this on the equal distribution of exercise groups between the BI and the MDI groups (Table 2), and the fact that the participants had equal pain improvements following the exercise programmes in the nested RCT [31].

The access to register data on RTW allowed for 100% follow-up on the primary outcome and the validity of DREAM has previously been demonstrated [41]. A fifth limitation was that appraisal of register data revealed minor inconsistencies at baseline between self-reported and register-based sick leave status. According to register data, 15 participants did not fulfil the inclusion criteria of sick leave ≥4 weeks. These participants were equally distributed between intervention groups and tentative per protocol analysis excluding these participants did not alter the results (adjusted HR = 0.70, 95% CI 0.44–1.12). It cannot be ruled out that the ITT analysis might introduce a minor degree of non-differentiated information bias. But this does not change the overall estimates of RTW and apart from maintaining the strength of randomization, the ITT analysis also displays high external validity since self-reported sick leave status is the only accessible information on the day of inclusion.

Sixth, the time spent on the MDI warrants consideration. Due to the setup of the study, participants in the MDI group waited 1–2 weeks after randomization before receiving the part of the intervention that differed from the BI group. Meanwhile, time at risk began at the day of randomization for both groups. Remembering the poor prognosis associated with prolonged sick leave [9–11, 15, 39] this was inexpedient but unfortunately unavoidable. Seventh, due to the sample size, there is approximately 30% risk of type 2 errors, i.e. a risk of overlooking an actual difference between the MDI and the BI intervention. We do not, however, consider power problems to explain the lack of difference, but rather characteristics of the population and intervention as described above.

Finally, only a minority of participants in the MDI experienced workplace involvement. In the latest review on workplace interventions, Cullen et al. present strong evidence on the positive work outcomes when applying multidomain interventions orchestrated from the workplace [17] and it could be argued that workplace involvement should have been mandatory. As previously described, this was not possible, because the majority of participants preferred to keep their health problems secret to their employers. As described, this discretion regarding health issues is rooted the Danish Health Information Law [25]. Whether a stronger focus on workplace involvement could have improved the results in the MDI group cannot be ruled out.

On the macro level, the “economic climate” is known to potentially affect sickness absence [7]. Our choice of outcome measure was constricted to four consecutive weeks of self-support, alternatively four consecutive weeks of holding a job supported by the social system. But since the study was performed during a period of economic recession in Denmark, exploratory analyses were performed allowing for the outcome RTW to be also 4 weeks of unemployment benefits and State Education Fund Grants (both reflecting readiness to return to work). These analyses still did not show significant differences in RTW between the groups but increased the HR in favor of the MDI (data not shown). Rather than interpreting the increased HR as the results of a successful MDI intervention, this merely reflects the termination of employment for some of the MDI participants. The combination of general economic recession and an intervention lasting several weeks may have contributed to the loss of jobs for some of the MDI participants.

In conclusion, no difference was found in RTW rates between the BI and the MDI group. Nor were there any differences in follow-up pain and disability between the groups. We do however assume that the evidence on the effect of multidisciplinary interventions in LBP [12, 14] and other musculoskeletal disorders [15, 17] is transferable to neck and shoulder pain. For clinical practice, several studies over the years e.g., [27, 40, 45–47] have suggested efficacy of a brief clinical intervention based on a non-injury approach with a focus of diminishing fear and restoring/maintaining normal daily activities. Add-on of a multidisciplinary intervention including a case manager as in the current study does not seem to improve RTW outcomes. Rather, evidence suggests the necessary involvement of workplaces.

Another implication for clinical practice derives from the above recognition: There is not only a need for efficient RTW interventions but also for increased focus on preventing sickness absence, i.e. how do clinicians identify patients at high risk of sickness absence? Feleus et al. recently published a study identifying three different trajectories for sickness absence (low, intermediate and high risk) in patients presenting in primary care with complaints of the arm, neck and shoulder [50]. They also identified bio-psycho-social variables associated with these trajectories. For whiplash-associated disorders, a tool predicting both chronic disability and full recovery has been developed [51, 52]. For neck pain however, current evidence does not support clinical use of neither prognostic nor prescriptive clinical prediction rules [53].

Better understanding of the prognostic factors and development of clinical prediction rules regarding RTW outcomes in neck and shoulder pain are suggested as future focus areas in research.
Acknowledgements  The authors thank all participants and stakeholders for their cooperation.

Funding  The Danish Rheumatism Foundation, Helga og Peter Korn- ing Foundation, Aase og Ejnar Danielsen Foundation, Aarhus University, Tryg Foundation.

Author Contributions  LTM: data management, analysis of data, primary responsible for the manuscript. CMS: assistance with analysis of data, critical appraisal of the manuscript. BS and ML: critical appraisal of the manuscript. DH: planning of the study, providing the clinical intervention (physiotherapist), critical appraisal of the manuscript. OKJ: planning of the study, providing the clinical intervention (physician), critical appraisal of the manuscript, power calculation. CVN: planning of the study, critical appraisal of the manuscript.

Compliance with Ethical Standards

Conflict of interest  Line Thorndal Moll, Christina Malmose Sta- pelfeldt, Berit Schütte-Christensen, David Høyrup-Christiansen, Ole Kudsk Jensen, Claus Vinther Nielsen, Merete Labriola declare that they have no conflict of interests.

Ethical Approval  All procedures followed were in accordance with the ethical standards of the regional ethical committee and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants included in the study.

Open Access  This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://crea tivecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

24. Faber E, Kuiper JJ, Burdorf A, Miedema HS, Verhaar JA. Treatment of impingement syndrome: a systematic review of the effects...
Degenerative findings on MRI of the cervical spine: an inter- and intra-rater reliability study

Line Thorndal Moll1,2,3*, Morten Wasmod Kindt4, Christina Malmose Stapelfeldt1,2 and Tue Secher Jensen4,5

Abstract

Background: Knowledge about the assessment reliability of common cervical spine changes is a prerequisite for precise and consistent communication about Magnetic Resonance Imaging (MRI) findings. The purpose of this study was to determine the inter- and intra-rater reliability of degenerative findings when assessing cervical spine MRI.

Methods: Fifty cervical spine MRIs from subjects with neck pain were used. A radiologist, a chiropractor and a second-year resident of rheumatology independently assessed kyphosis, disc height, disc contour, vertebral endplate signal changes, spinal canal stenosis, neural foraminal stenosis, and osteoarthritis of the uncovertebral and zygapophyseal joints. An evaluation manual was composed containing classifications and illustrative examples, and ten of the MRIs were evaluated twice followed by consensus meetings to refine the classifications. Next, the three readers independently assessed the full sample. Reliability measures were reported using prevalence estimates and unweighted kappa (κ) statistics.

Results: The overall inter-rater reliability was substantial (κ ≥ 0.61) for the majority of variables and moderate only for zygapophyseal osteoarthritis (κ = 0.56). Intra-rater reliability estimates were higher for all findings.

Conclusions: The present classifications for some of the most common cervical degenerative findings yielded mainly substantial inter-rater reliability estimates and substantial to almost perfect intra-rater reliability estimates.

Trial registration: Regional Data Protection Agency (J.no. 1–16–02-86-16). The letter of exemption from the Regional Ethical Committee is available from the author on request (case no. 86 / 2017).

Keywords: Magnetic resonance imaging, Reliability, Cervical spine, Degenerative, Classification, MRI, Agreement

Background

Although not recommended as routine imaging in neck pain [1, 2], the number of cervical MRIs has increased by 18% compared to a 4.5% increase in neck pain prevalence over recent years in Denmark [2–4]. While patients believe in MRI to unveil the true cause of their pain [5], health care professionals appreciate the advantages of MRI compared with other modalities of diagnostic imaging. The non-invasiveness, absence of radiation exposure and the capacity to discriminate soft tissue changes are all highly valued in the field of musculoskeletal imaging.

When communicating MRI findings, the importance of consistency and precision remains unaltered. Both for academic and clinical purposes, a prerequisite for such consistency and precision is reliability in MRI assessments. Reliability is defined as “the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions” [6]. In the case of MRI, this means that while the images do not change, reliability reflects whether the image interpretation remains the same when assessed by different raters (inter-rater reliability) or by the same rater at different times (intra-rater reliability).

Previous reliability studies on cervical spine MRI have found moderate to almost perfect inter-rater reliability changes are highly valued in the field of musculoskeletal imaging.

© The Author(s). 2018 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
in the assessments of disc-related parameters (kappa (K) 0.44\textsuperscript{[7]}, K 0.43–0.65 \textsuperscript{[8]} and K 0.73–0.83 \textsuperscript{[9]}). Almost perfect reliability has been reported for assessments of neural foraminal stenosis (K > 0.9 \textsuperscript{[10]}), fair reliability for facet joint arthrosis (K 0.23–0.38 \textsuperscript{[11]}), and moderate to substantial reliability for spinal canal stenosis (K 0.55–0.72 \textsuperscript{[11]}). Most studies have focused on only one or a few degenerative variables \textsuperscript{[7–13]} and compared readers with similar educational backgrounds and levels of experience \textsuperscript{[7–10, 12–14]}.

To our knowledge, only one reliability study on cervical spine MRI has covered a broad range of common degenerative findings \textsuperscript{[14]} for which reason, further studies are needed.

**Objective**

To determine the inter- and intra-rater assessment reliability of degenerative findings (kyphosis, disc height, disc contour, vertebral endplate signal changes, spinal canal stenosis, neural foraminal stenosis, uncovertebral osteoarthritis and zygapophyseal osteoarthritis) on MRI of the cervical spine.

**Methods**

**Subjects**

Fifty MRIs of the cervical spine were chosen from among subjects previously enrolled in a randomized controlled trial (RCT) \textsuperscript{[15]}. Subjects for the RCT were recruited from primary health care professionals (physiotherapists, chiropractors and general practitioners (GPs)). If subjects fulfilled the inclusion criteria (age 18–60 years, part-time or full-time sick leave for 4–16 weeks owing to neck pain or shoulder pain, and fluency in Danish), their GPs referred them to The Spine Centre, Silkeborg Regional Hospital, Denmark. For the current study, the predefined inclusion criterion was the availability of a cervical spine MRI with a satisfactory signal-to-noise ratio. After assessment by the most experienced reader, 32 MRIs were excluded based on unsatisfactory signal-to-noise ratio. By choosing every second MRI among those remaining, 50 MRIs were selected for the current study. A study flow-chart is seen in Fig. 1.

**Data collection - images**

The MRIs were provided from five different hospitals collaborating with The Spine Centre. The majority of the images were obtained using a 1.5 T field strength. All MRIs comprised sagittal T1-weighted and T2-weighted sequences, while an axial T2 sequence was available for 94% and oblique T2 sequences were available for 82% of the images.

**Data collection – readers**

The three readers (Readers A, B and C) all assessed the images independently over a time frame of 5–8 weeks. Reader A was a second-year resident of rheumatology with no previous formal education in MRI assessment. She had 9 years of postgraduate clinical experience including assessment of spinal MRI for clinical purposes. Reader B was an experienced radiologist having worked with musculoskeletal MRI for 25 years, mostly on a daily basis. Reader C was a chiropractor who had completed a 1-year fulltime internship in spinal MRI in a radiology department. He had another 10 years of clinical and academic experience with spinal MRI. Prior to the study, Reader B taught Reader A assessment of cervical spine MRI for 2 h. Following this two-hour session, Reader A completed 50 clinical narrative reports of cervical spine MRIs from patients with neck pain with or without radiculopathy. These were not part of the current study. The reports were corrected if necessary and approved by Reader B.

For the intra-rater reliability assessment, Reader A assessed all the images twice. The second assessment took place after 6 weeks to prevent recollection of the first assessments.

**Evaluation manual, piloting and workstations**

Based on the literature \textsuperscript{[10–14, 16–24]}, an evaluation manual with written and visual classifications of the findings was made by Reader A, adjusted and approved by Readers B and C. Next, 10 MRIs from the study sample were evaluated twice followed by consensus meetings. This piloting served the purpose of refining both the classifications in the evaluation manual and the practice of the readers. All images were de-identified, leaving
the readers blinded to demographic and clinical data as well as previous assessments. The images were assessed on radiological work stations using Vitrea Core (version 1.0.0.404, Vital Images Inc.).

**Variables**

Classifications for common and degenerative MRI findings were developed based primarily on the existing literature [10–13, 16–19, 23–26] and on experiences from the piloting. An effort was made to create definitions that were as simple as possible [14], assuming that simplicity is essential for clinical applicability. The most common degenerative findings were chosen, including kyphosis and vertebral endplate signal changes; all are routinely considered by radiologists assessing cervical spine MRIs at Silkeborg Regional Hospital. All the classifications yielded categorical (but not ordinal) data. The complete list of variables is presented in Table 1. Except for kyphosis, these findings were assessed for each of the six cervical disc levels (level C2/C3 to C7/T1). Furthermore, the neural foramina, uncovertebral and zygapophyseal joints were assessed separately on the left and right hand side. The evaluation manual is available in Additional file 1.

### Table 1 MRI findings and corresponding classifications

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis</td>
<td>0</td>
<td>Normal or straightened lordosis</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Kyphosis</td>
</tr>
<tr>
<td>Disc height</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Reduced height</td>
</tr>
<tr>
<td>Disc contour</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Bulge or protrusion</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Extrusion</td>
</tr>
<tr>
<td>Spinal canal stenosis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>&gt; 50% obliteration of CSF, no cord deformity</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>&gt; 50% obliteration of CSF with cord deformity but no signal change</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>&gt; 50% obliteration of CSF with cord deformity and signal change</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Type 1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Type 2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Type 3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Mixed type 1 and 2</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Definite osteoarthritis</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis</td>
<td>0</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Definite osteoarthritis</td>
</tr>
<tr>
<td>Neural foraminal stenosis</td>
<td>0</td>
<td>Normal or &lt; 50% fat obliteration</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>≥ 50% fat obliteration with or without morphological changes of the nerve root</td>
</tr>
</tbody>
</table>

**Data entry and statistical analysis**

All three readers independently entered and stored data using Epidata (Version 3.1., The EpiData Association, Odense, Denmark, 2003–2004). If assessment of a certain finding was not possible due to the available sequences, the particular finding was allotted the value ‘9’ representing ‘missing’.

In accordance with the recommendations for reliability studies [27], 50 MRIs were included in the current study. Prior to the kappa (K) calculations, all readers’ prevalence assessments were calculated, one variable at a time. This tabulation of data offered the opportunity of 1) assessing the sample homogeneity and 2) identifying any possible systematic differences between the readers; as both can affect the K estimates [27, 28]. Tabulation thus allowed for a clearer impression of agreement and possible misclassification than offered by the K value alone. Tabulation also provided estimates for observed agreement (OA) and agreement by chance (AC) for the pairwise analyses. For the overall three-reader analysis, OA was calculated by computing the number of observations with complete agreement and dividing this number with the total number of anatomical sites assessed. The three-reader AC was calculated by multiplication of...
the marginal fractions [27]. Reliability measures were computed using unweighted kappa statistics owing to the categorical (as opposed to ordinal) nature of the data. Given the condition of total independence among the readers, K is defined as

\[ K = \frac{OA - AC}{1 - AC} \]

where OA is observed agreement and AC agreement by chance [29]. Reliability measures were computed for the readers in pairs (A1B1, A1C1, B1C1, A1A2) and over-all (A1B1C1). Acknowledging the influence of prevalence on the K estimates [27, 28], these were only computed whenever the readers in question agreed on prevalences ≥10%. For each disc level, the left and right hand side assessments of neural foraminal stenosis, uncovertebral and zygapophyseal osteoarthritis were pooled before computing reliability estimates. The interpretation of K values followed the suggestions by Landis & Koch [29]:

<table>
<thead>
<tr>
<th>K value</th>
<th>Strength of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0</td>
<td>Poor</td>
</tr>
<tr>
<td>0.0–0.2</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21–0.4</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41–0.6</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61–0.8</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81–1.0</td>
<td>Almost perfect</td>
</tr>
</tbody>
</table>

K values were reported using 95% confidence intervals and additional information on OA and AC were supplied for all findings. Analyses were performed using the STATA (version 15.0; Stata Corporation, College Station, Texas, USA) software package.

**Ethics**

All subjects provided written informed consent. The study was approved by the Regional Data Protection Agency (J.no. 1–16–02-86-16). Approval by the regional ethical committee was not needed due to the study’s methodological nature. The letter of exemption from The Central Denmark Region Committees on Health Research Ethics is available from the author on request (case no. 86 / 2017).

**Results**

The majority of the subjects were female (n = 31; 62%) with a mean age of 43.7 years (SD = 9.2). The prevalence of positive findings for all readers can be seen in Additional file 2. For vertebral endplate signal changes, prevalence estimates were below 10% and thus too low for K statistics. For the remaining degenerative findings, prevalence estimates allowed for kappa statistics including one to six anatomical sites (e.g. 2 disc levels ~ 100 observations included in K analysis for spinal canal stenosis). Further scrutiny of the prevalence table revealed a slight tendency for Reader C to assign the label “reduced disc height” more frequently. Otherwise no systematic differences among the readers were identified.

As shown in Table 2, the overall inter-rater reliability (A1B1C1) ranged from moderate to almost perfect for the majority of the findings (substantial to almost perfect for kyphosis and neural foraminal stenosis; moderate to almost perfect for spinal canal stenosis; and moderate to substantial for disc height, disc contour, uncovertebral and zygapophyseal osteoarthritis). Exploratory analyses were made to assess the inter-rater reliability of neural foraminal stenosis when including only MRIs with oblique images (Additional file 3). This did not change the reliability estimates but broadened the confidence intervals slightly.

The intra-rater reliability estimates (Table 3) were slightly better than those for inter-rater reliability. Almost perfect reliability was found for kyphosis and substantial to almost perfect reliability for disc contour, uncovertebral osteoarthritis and neural foraminal stenosis. For spinal canal stenosis and zygapophyseal osteoarthritis, moderate to almost perfect intra-rater reliability was found while moderate to substantial reliability was found for disc height.

**Discussion**

To our knowledge, this is the first reliability study covering eight common cervical MRI findings. The overall inter-rater reliability was substantial for all variables except zygapophyseal osteoarthritis where moderate reliability was found. Intra-rater reliability was substantial for the majority of variables and almost perfect for kyphosis. These reliability estimates reflect that the observed agreement notably exceeds the agreement that can be expected by chance.

For disc degeneration, other studies [9, 12] reported higher reliability estimates than the disc height estimates in the current study. Although the use of intraclass correlation coefficient in the study by Jacobs et al. [12] does not allow for direct comparison, possible explanations for the reliability differences are the use of a ubiquitously accessible reference image of a normal disc [12] and the notable experience among readers with the same educational background [9].

For disc contour, the reliability estimates were similar to those of other studies despite the fact that we used a three-category classification compared to the previously reported dichotomous classifications [8, 30, 31] and comparison of more experienced readers [30, 31].

For spinal canal stenosis, the current study’s unweighted reliability estimates exceeded those previously reported by use of weighted kappa statistics [13, 32], although the use of weights are expected to yield higher estimates. A higher
<table>
<thead>
<tr>
<th>MRI finding</th>
<th>n</th>
<th>Reader pair</th>
<th>Observed agreement (%)</th>
<th>Agreement by chance (%)</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50</td>
<td>A1B1</td>
<td>92.0</td>
<td>56.4</td>
<td>0.82 (0.75; 0.89)</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>A1C1</td>
<td>89.8</td>
<td>53.6</td>
<td>0.78 (0.71; 0.85)</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>B1C1</td>
<td>89.8</td>
<td>52.8</td>
<td>0.78 (0.71; 0.86)</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>A1B1C1</td>
<td>85.7</td>
<td>31.2</td>
<td>0.79 (0.73; 0.85)</td>
</tr>
<tr>
<td>Disc height&lt;sup&gt;b&lt;/sup&gt;</td>
<td>150</td>
<td>A1B1</td>
<td>92.0</td>
<td>52.8</td>
<td>0.83 (0.74; 0.92)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>A1C1</td>
<td>80.0</td>
<td>52.8</td>
<td>0.58 (0.46; 0.69)</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>B1C1</td>
<td>77.3</td>
<td>50.0</td>
<td>0.55 (0.42; 0.68)</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>A1B1C1</td>
<td>74.7</td>
<td>26.4</td>
<td>0.65 (0.57; 0.74)</td>
</tr>
<tr>
<td>Disc contour&lt;sup&gt;b&lt;/sup&gt;</td>
<td>177</td>
<td>A1B1</td>
<td>76.8</td>
<td>43.4</td>
<td>0.59 (0.49; 0.70)</td>
</tr>
<tr>
<td></td>
<td>177</td>
<td>A1C1</td>
<td>79.7</td>
<td>43.3</td>
<td>0.64 (0.53; 0.74)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>B1C1</td>
<td>80.0</td>
<td>47.6</td>
<td>0.62 (0.52; 0.72)</td>
</tr>
<tr>
<td></td>
<td>177</td>
<td>A1B1C1</td>
<td>68.4</td>
<td>21.7</td>
<td>0.61 (0.54; 0.69)</td>
</tr>
<tr>
<td>Spinal canal stenosis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>100</td>
<td>A1B1</td>
<td>97.0</td>
<td>76.0</td>
<td>0.88 (0.68; 1.00)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>A1C1</td>
<td>91.0</td>
<td>73.5</td>
<td>0.66 (0.47; 0.83)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>B1C1</td>
<td>92.0</td>
<td>74.3</td>
<td>0.69 (0.48; 0.86)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>A1B1C1</td>
<td>90.0</td>
<td>63.0</td>
<td>0.74 (0.57; 0.86)</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>222</td>
<td>A1B1</td>
<td>90.1</td>
<td>68.0</td>
<td>0.69 (0.57; 0.81)</td>
</tr>
<tr>
<td></td>
<td>237</td>
<td>A1C1</td>
<td>89.0</td>
<td>68.6</td>
<td>0.65 (0.53; 0.77)</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>B1C1</td>
<td>87.4</td>
<td>70.9</td>
<td>0.57 (0.43; 0.71)</td>
</tr>
<tr>
<td></td>
<td>222</td>
<td>A1B1C1</td>
<td>83.3</td>
<td>53.0</td>
<td>0.65 (0.51; 0.76)</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>270</td>
<td>A1B1</td>
<td>94.8</td>
<td>74.2</td>
<td>0.80 (0.70; 0.90)</td>
</tr>
<tr>
<td></td>
<td>144</td>
<td>A1C1</td>
<td>87.5</td>
<td>74.9</td>
<td>0.50 (0.31; 0.70)</td>
</tr>
<tr>
<td></td>
<td>184</td>
<td>B1C1</td>
<td>85.9</td>
<td>78.9</td>
<td>0.33 (0.13; 0.53)</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>A1B1C1</td>
<td>83.0</td>
<td>61.0</td>
<td>0.56 (0.43; 0.70)</td>
</tr>
<tr>
<td>Neural foraminal stenosis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>268</td>
<td>A1B1</td>
<td>90.7</td>
<td>64.1</td>
<td>0.74 (0.65; 0.84)</td>
</tr>
<tr>
<td></td>
<td>287</td>
<td>A1C1</td>
<td>90.2</td>
<td>64.2</td>
<td>0.73 (0.63; 0.82)</td>
</tr>
<tr>
<td></td>
<td>275</td>
<td>B1C1</td>
<td>87.6</td>
<td>65.8</td>
<td>0.64 (0.53; 0.75)</td>
</tr>
<tr>
<td></td>
<td>268</td>
<td>A1B1C1</td>
<td>84.0</td>
<td>46.0</td>
<td>0.73 (0.63; 0.82)</td>
</tr>
</tbody>
</table>

<sup>a</sup>n refers to the number of MRIs assessed  
<sup>b</sup>n refers to the number of disc levels assessed  
<sup>c</sup>n refers to the number of anatomical sites assessed (by pooling right and left hand side)  

Table 3 Intra-rater reliability estimates

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>n</th>
<th>Reader pair</th>
<th>Observed agreement (%)</th>
<th>Agreement by chance (%)</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50</td>
<td>A1A2</td>
<td>96.0</td>
<td>59.6</td>
<td>0.90 (0.85; 0.96)</td>
</tr>
<tr>
<td>Disc height&lt;sup&gt;b&lt;/sup&gt;</td>
<td>200</td>
<td>A1A2</td>
<td>84.0</td>
<td>51.5</td>
<td>0.67 (0.57; 0.77)</td>
</tr>
<tr>
<td>Disc contour&lt;sup&gt;b&lt;/sup&gt;</td>
<td>174</td>
<td>A1A2</td>
<td>88.5</td>
<td>43.9</td>
<td>0.80 (0.71; 0.87)</td>
</tr>
<tr>
<td>Spinal canal stenosis&lt;sup&gt;b&lt;/sup&gt;</td>
<td>50</td>
<td>A1A2</td>
<td>94.0</td>
<td>76.6</td>
<td>0.73 (0.51; 0.90)</td>
</tr>
<tr>
<td>Vertebral endplate signal change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>281</td>
<td>A1A2</td>
<td>90.4</td>
<td>67.0</td>
<td>0.71 (0.61; 0.81)</td>
</tr>
<tr>
<td>Zygapophyseal osteoarthritis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>240</td>
<td>A1A2</td>
<td>90.8</td>
<td>68.8</td>
<td>0.71 (0.59; 0.82)</td>
</tr>
<tr>
<td>Neural foraminal stenosis&lt;sup&gt;c&lt;/sup&gt;</td>
<td>287</td>
<td>A1A2</td>
<td>90.6</td>
<td>62.6</td>
<td>0.75 (0.66; 0.84)</td>
</tr>
</tbody>
</table>

<sup>a</sup>n refers to the number of MRIs assessed  
<sup>b</sup>n refers to the number of disc levels assessed  
<sup>c</sup>n refers to the number of anatomical sites assessed (by pooling right and left hand side)
number of readers (six [13] and nine [32]) could explain this difference, but even when compared to the three most experienced readers in these studies, better reliability estimates were still achieved in the current study. The most probable reason appears to be the limited introduction of their classification [13, 32]. When using both written and visual descriptions, our moderate to almost perfect reliability among readers with considerable experience differences suggest good applicability of this classification of spinal canal stenosis.

For zygapophysial osteoarthritis, both the intra- and inter-rater reliability estimates were better than previously reported [11], which is most likely explained by the use of a dichotomous variable in the current study compared to a classification with four severity categories [11].

For neural foraminal stenosis, this study still achieved higher reliability estimates compared to studies with more experienced readers [30, 31]. The inferior reliability estimates may be explained by unclear definitions [30] and by low prevalence estimates together with images obtained using a 0.5 T field strength [31]. Compared to the study from which we modified the classification of neural foraminal stenosis [10], the current study was unable to reach the same almost perfect reliability estimates (K > 0.9). Nevertheless, we consider the substantial to almost perfect reliability to be satisfactory, bearing in mind differences in reader experience and the heterogeneous image material (i.e. images with different field strengths and available sequences). The modified classification (dichotomous versus the original four categories) proved reliable and the association with clinical findings has previously been reported [33].

Methodological considerations
A limitation of the study is that it was not preceded by a power calculation. However, the confidence intervals for the K estimates only comprised more than two levels (e.g. from moderate to almost perfect for spinal canal stenosis) in a minority of cases. A larger sample would have narrowed the confidence intervals but would probably not have caused substantial changes in the reliability estimates.

Another limitation is the involvement of only reader A in the intra-rater analysis. Two considerations explain this: 1) previous reliability studies found higher [7–9, 12, 14, 21] or similar/higher [10, 11, 13] intra-rater reliability than inter-rater reliability and 2) involvement of reader A was necessary since a future prognostic study will involve MRI assessments performed by reader A. As for the inter-rater reliability, the study included three readers, only one of these being a radiologist. However, the results suggest that our method is applicable among other health care professionals (i.e. rheumatologists and chiropractors) in a controlled research setting. Involvement of other relevant healthcare professionals, e.g. spine surgeons, would have been desirable but was unfortunately not possible.

Owing to the properties of K, the measure does not disentangle systematic and random misclassification [28]. Therefore, we provided the prevalence tables from which we find no suspicion of systematic misclassification.

The prevalence table discloses a notable difference in the number of disc levels assessed for disc contour on levels C2/C3, C3/C4 and C7/T1: Reader A assessed fewer levels than Readers B and C owing to the lack of axial images of the selfsame disc levels. This discrepancy suggests a difference among the readers, and whether this partly explains why higher reliability estimates were not achieved for disc contour cannot be refuted.

Another potential limitation is that all MRIs were derived only from individuals with neck pain. But since cervical spine MRI is seldom performed in patients without neck pain and since the future use of the evaluation manual applies to patients with neck pain, we consider the current sample appropriate for its purpose.

Finally, a potential limitation of the study is the heterogeneous image material (MRIs were performed at five different hospitals. Different field strengths and sequences were available). Yet, as it resembles everyday clinical practice, this was an intended challenge and an attempt was made to manage this heterogeneity by using a standardized evaluation manual. The differences between OA and AC (Tables 2 and 3) reflect that both inter- and intra-rater agreement notably exceed the agreement that can be expected by chance. Furthermore, the high levels of observed agreement reflect only a minor degree of misclassification. Based on these observations of OA, our interpretation is that the evaluation manual and the standardized procedures explain the high levels of agreement rather than pure chance when assessing heterogeneous images.

Ultimately, the heterogeneous image material and the use of three different health care professionals both add to the generalizability and thus constitute strengths of the study. The blinding of the readers, the use of simple and easily comprehensible classifications along with regular encouragement to follow the evaluation manual, are other important strengths of the study.

In contrast to the controlled settings of the current study, a study comparing narrative MRI reports demonstrated considerable variability [34]. In this study [34], a patient with low back pain and right L5 radicular symptoms had lumbar spine MRI performed at 10 different MRI centers within 3 weeks. Comparison of the 10 narrative reports revealed considerable variability; none of the 49 described findings occurred in all 10 reports and only one finding occurred in nine reports. Even if this amount of variability is unusually large [34], it supports our clinical experience that variability also prevails in the interpretation of cervical...
spine MRIs. A possible way to overcome this is by using classifications sufficiently comprehensible to be applied 1) by different health care professionals and 2) when assessing heterogeneous images from different MRI scanners. Such classifications were presented in the current study. Confirmatory studies will be needed. If those studies were to involve experienced radiologists, provide proper training for lesser experienced MRI readers, and use an evaluation manual, better reliability might be achieved in clinical settings. So far, the results suggest that the evaluation of MRI findings can be used in controlled research settings studying individuals with neck pain. Suggestions for future research include comparison of reliability with and without the use of an evaluation manual. Also, including more than one of each health care professional could allow for comparison of experience levels both among and within different types of health care professionals.

Conclusions
In conclusion, the current study found substantial reliability for the majority of included MRI findings. This suggests that the present classifications are sufficiently comprehensible to be applied by different health care professionals when assessing images from different MRI scanners. In our view, the proposed classifications are sufficiently reliable to be used for both quality assurance and further research purposes.

Additional files

- **Additional file 1**: The evaluation manual used for assessment of the MRIs. (DOCX 2347 kb)
- **Additional file 2**: A prevalence table reporting the frequency of positive findings for all the readers. (DOCX 30 kb)
- **Additional file 3**: A table of sensitivity analyses. For neural foraminal stenosis, kappa estimates are presented comparing the assessments of all images vs. only images with available oblique slices. (DOCX 16 kb)

Abbreviations
AC: Agreement by chance; CSF: Cerebrospinal fluid; GP: General practitioner; MRI: Magnetic resonance imaging; OA: Observed agreement; RCT: Randomized controlled trial; SD: Standard deviation; K: Kappa

Acknowledgements
A special thanks to Brian Højgaard for readily providing technical support whenever needed.

Funding
This work was supported by the Tryg Foundation, Aarhus University Denmark, Danish Rheumatism Association, and Aase and Ejnar Danielsen Foundation.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors’ contributions
LTM, MWK and TSJ designed the study and collected the data. LTM performed the statistical analyses and drafted the manuscript. All the authors contributed to the interpretation of data. All the authors critically revised and approved the final manuscript.

Ethics approval and consent to participate
Written informed consent was provided from the participants. The study was approved by the Regional Data Protection Agency (J.no. 1–16–02-86-16). Approval by the Regional Ethical Committee was not needed due to the study’s methodological nature. The letter of exemption from the Regional Ethical Committee is available from the author on request (case no. 86 / 2017).

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1DEFACTUM, Central Denmark Region, P.P. Oerums Gade 11, bygn. 1B, DK-8000 Aarhus C, Denmark. 2Section of Clinical Social Medicine and Rehabilitation, Department of Public Health, Aarhus University, P.P. Oerums Gade 9-11, bygn. 1B, DK-8000 Aarhus C, Denmark. 3Spine Centre, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways, Silkeborg Regional Hospital, Falkevej 1-3, DK-8600 Silkeborg, Denmark. 4Department for Diagnostic Imaging, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways, Silkeborg Regional Hospital, Falkevej 1-3, DK-8600 Silkeborg, Denmark. 5Nordic Institute of Chiropractic and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark.

Received: 27 March 2018 Accepted: 13 August 2018
Published online: 16 October 2018

References


Study 3
Title
Prediction of work participation within 2 years in sickness absentees with neck or shoulder pain: the contribution of demographic, patient-reported, clinical and imaging information

Journal
Journal of Occupational Rehabilitation

The names of the authors
Line Thorndal Moll\textsuperscript{1,2,3}, Anne Mette Schmidt\textsuperscript{1,2,4}, Christina Malmose Stapelfeldt\textsuperscript{1,2}, Merete Labriola\textsuperscript{1,2}, Ole Kudsk Jensen\textsuperscript{3}, Morten Wasmod Kindt\textsuperscript{5}, Tue Secher Jensen\textsuperscript{5,6}, Berit Schiøttz-Christensen\textsuperscript{7}

Corresponding author's affiliation and e-mail
\textsuperscript{1} DEFACTUM, Central Denmark Region, Aarhus, Denmark
P.P. Oerums Gade 11, bygn. 1B
DK-8000 Aarhus C.

\textsuperscript{2} Section of Clinical Social Medicine and Rehabilitation, Department of Public Health, Aarhus University, Denmark
P.P. Oerums Gade 9-11, bygn. 1B
DK-8000 Aarhus C.

\textsuperscript{3} Spine Centre, Diagnostic Centre, Silkeborg Regional Hospital, Denmark
Falkevej 1-3
DK-8600 Silkeborg

\textit{e-mail:} linethor@rm.dk

Abstract

Purpose
This cohort study explored the contribution of demographic, patient-reported, clinical, and magnetic resonance imaging (MRI) information to the prediction of work participation in sickness absentees with neck or shoulder pain.
Methods

From a secondary care setting, 168 sickness absentees with neck or shoulder pain were included. Based on registry data, a successful work outcome was defined as ≥ 75% work participation score (WPS) from Weeks 30 to 104 after enrolment. Prognostic variables were categorized into four domains (demographic, patient-reported, clinical, and MRI) resembling the order of information obtained in a clinical setting. Crude logistic regression analyses were used to identify prognostic variables for each domain (p<0.2). This was followed by multivariable analyses including the identified variables in a domain-wise order. For each added domain, the probability of successful WPS was dichotomized leaving two possible classifications: ≥ 50% chance of successful WPS or not. In cross-tabulations of chance and the actual WPS outcome, positive and negative predictive values (PPV and NPV), sensitivity, specificity and area under the curve (AUC) were calculated.

Results

The combination of demographic and patient-reported variables yielded an NPV of 70% and a PPV of 72%, while specificity was 68%, sensitivity 74% and AUC 0.76. None of these values improved notably by adding clinical and MRI variables as predictors of successful WPS.

Conclusions

These results suggest that - among sickness absentees with neck or shoulder pain – clinical and MRI variables provide no additional information for the prediction of work participation compared with only demographic and patient-reported information.

Keywords

Prognosis, sick leave, neck pain, shoulder pain, magnetic resonance imaging.

Background

During sick leave, estimating work prognosis is important as it helps guide expectations for all involved stakeholders. The responsibility for this usually resides with healthcare professionals and it is often requested of them by authorities involved in return-to-work (RTW) coordination and the disbursement of social benefits, i.e.
sickness benefits. When accommodating this request of work prognosis assessment for patients with neck or shoulder pain, healthcare professionals have access to different domains of information. In a clinical setting, such information is typically obtained in the following order: 1) demographics are known already when scheduling the clinical encounter, and are followed by 2) patient-reported information which is a necessary prerequisite for performing 3) a focused clinical examination, and deciding whether 4) imaging is needed (e.g. magnetic resonance imaging (MRI)).

From this range of information, the literature offers insight into known prognostic factors. From the demographic domain, older age is negatively associated with RTW [1, 2], while for the impact of sex, results are conflicting [1-3]. Patient-reported information that is negatively associated with work outcomes includes symptom intensity, sick leave duration [1, 3] and higher physical work demands [1]. From the clinical domain, low back pain studies have found a negative association between radiating pain and RTW in the acute phase (< 6 weeks) [4] and moderate evidence for no association with RTW in the sub-acute and chronic phases [2]. Also from the clinical domain, fibromyalgia is associated with poorer work participation [5]. From the MRI domain, a single study involving sickness absentees with low back pain found Modic type 1 changes on MRI to predict unsuccessful RTW [6]. However, the prognostic value of imaging has been described as "a gap in the literature" [7]. This gap still deserves investigation as the number of MRIs has increased more than the prevalence of people suffering from neck and shoulder pain [8-10]. Moreover, the referral patterns have changed over recent years; e.g. in Denmark, not only hospital physicians but also primary care clinicians (physicians and chiropractors) can request MRI. For clinicians in both primary and secondary care, this change in MRI referral patterns has changed the range of information available when estimating work prognosis.

The literature thus offers insight into factors associated with work prognosis. But to our knowledge, no previous studies have examined to what extent the different domains of information contribute to the prediction of work prognosis. Thus healthcare professionals’ assessment of work prognosis may be based on gut-feeling rather than evidence. Therefore, the aim of this study was to explore the degree to which demographic, patient-reported, clinical, and MRI information contribute to the prediction of work participation in sickness absentees with neck or shoulder pain.
Methods

Design and participants

Using a cohort design, 168 sickness absentees referred to secondary care for assessment of neck or shoulder pain were included in this study. They were all part of a randomized controlled trial (RCT) [11] that took place from 2009-2014 and compared the work outcomes after two different hospital-based interventions (multidisciplinary vs. brief intervention). The inclusion criteria for the RCT were: sick leave 1-4 months owing to neck or shoulder pain, age 18-60 years, and fluency in Danish. Exclusion criteria were: alcohol or drug addiction, pregnancy, specific musculoskeletal/neurological disease, surgery within the past year, plans for surgery in the future, and primary psychiatric disorder. Further details regarding recruitment and interventions have been described previously [11].

Context (Danish legislation on sickness benefits)

According to The Danish Sickness Benefit Act, sickness absentees are entitled to financial compensation during sick leave. This compensation can amount to sickness benefits or full wages depending on the union award conditions the worker is entitled to. At the time of the RCT (2009-2014), the employer covered these expenses for the first 21-30 days of a sick leave spell (called the employer paid period) after which the municipality reimbursed some of the employer's expenses (the sickness benefit amount) until RTW was achieved. In Denmark, a special agreement (§56) can be settled upon in cases of recurrent short-term sick leave; it entitles the employer to be reimbursed from the first day of sick leave, thereby reducing the employer's expenses [12].

Outcome: Work participation score (WPS)

Data on work outcomes were supplied by the Danish Register for Evaluation of Marginalization (DREAM) which covers the source of income (financial self-support or public transfer benefits) on a weekly basis since July 1991. The registrations are ordered hierarchically; if different transfer benefit codes are registered in the same week, the highest-ranking code will overwrite the others. Sickness benefits have a high priority and are only overwritten in cases of emigration, age-related pensions or death [13].

The work participation score (WPS) is a fraction yielding scores between 0 and 100% [14]. The numerator comprises the number of weeks with financial self-support (interpreted as working weeks) as well as weeks with state education fund grants and related benefits (benefits granted in cases of re-education). The denominator
comprises the total number of follow-up weeks [14]. In this study, three individuals were granted early retirement. For these three, the denominator comprised the number of follow-up weeks prior to the first appearance of retirement benefits in DREAM. Since the median time until the first RTW (4 consecutive weeks of financial self-support) was 29 weeks, follow-up time comprised the Weeks 30-104 and a successful outcome was defined as a WPS ≥ 75%. The 75% threshold was chosen because it has shown moderate to substantial agreement compared with other RTW measures [14] and because it defined an improvement compared with the mean WPS the year before enrolment (74%). The following denotations were used: successful WPS (s-WPS) ≥ 75% and unsuccessful WPS (u-WPS) < 75%.

**Prognostic variables: Demographic and patient-reported domain**

These data were collected from questionnaires filled in by the participants prior to their first meeting with the health care professionals in the RCT [11]. The information provided covered demographics (sex and age) and the following patient-reported measures: pain intensity (11-point numeric rating scale) [15], whiplash trauma (yes/no) [16], education (≥ 3 years, yes/no), worker’s compensation claim (yes/no) [17], and Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ) [18]. ÖMPSQ was originally developed as a screening tool to identify psychosocial risk factors for prolonged disability and sick leave; the higher the score (maximum = 210), the higher the risk of prolonged disability.

**Prognostic variables: Clinical domain**

Data on radiculopathy and number of tender points [19, 20] stemmed from the clinical examination performed by the rheumatologist prior to enrolment in the RCT [11]. Radiculopathy describes a clinical assessment based on 1) pain radiating from the neck to the upper extremity and 2) one or more positive neurological signs: weakened deep tendon reflexes, decreased muscle strength, dermatomal sensory deficits or a positive foraminal compression test. The reliability of the rheumatologist's tender points count has previously been reported as excellent [21] and a threshold of ≥ 11 tender points was chosen, since this is still commonly used among clinicians to assess fibromyalgia [22]. In the clinical examination, the tender points count was used as a measure of sensitization of the nociceptive system [23]. Since data were not sufficient to support the diagnosis of fibromyalgia, this term was not used and a dichotomized threshold of the tender points count was used instead. The interventions in the original RCT [11] (multidisciplinary and brief intervention) were also included.
Prognostic variables: MRI domain

MRI of the cervical spine was ordered for all participants unless clinical examination revealed unambiguous signs of shoulder pathology (n = 7). For different reasons, MRI of the cervical spine was only available for 97 of the participants (no access to MRI: n = 32 and unsatisfactory signal-to-noise ratio assessed by a senior radiologist: n = 32 (unpublished). All MRIs were de-identified and assessed by the first author (physician with 9 years of post-graduate clinical experience) who was blinded to the study outcomes. The inter-rater reliability of the MRI assessments has previously been reported with un-weighted kappa values ranging from 0.56 to 0.79 (unpublished) while intra-rater reliability yielded kappa values ranging from 0.67 to 0.90 (unpublished). The MRI variables listed in Table 1 were evaluated at disc levels C2/C3 to C7/T1. For each participant, an MRI finding was defined as positive if the finding was present at ≥ 1 disc level, i.e. two levels with disc height reduction were counted only once.

Statistical Analyses

Crude logistic regression analyses were performed for all baseline variables hypothesized to affect WPS, and variables yielding p-values < 0.2 were carried forward as has been done in other studies [24, 25]. In the multivariable analyses, the prognostic variables were included using a domain-wise approach resembling the typical order of information obtained in a clinical setting. First, demographic variables were included, second patient-reported, third clinical, and finally MRI variables – these are referred to as Models 1, 2, 3 and 4. For each model, the probability of s-WPS was calculated and dichotomized, i.e. the participants were classified as having ≥ 50% chance of s-WPS or not. By comparing the estimated chance of s-WPS with whether s-WPS actually happened or not, sensitivity, specificity, positive and negative predictive values (PPV and NPV), and area under the curve (AUC) were calculated. In this study, sensitivity denotes the proportion of u-WPS correctly classified as '<50% chance of s-WPS', and specificity, the proportion of s-WPS that is correctly classified as '≥50% chance of s-WPS'. In accordance with this, NPV denotes the proportion of sickness absentees classified as '≥50% s-WPS chance' who actually achieved s-WPS. PPV denotes the proportion of sickness absentees classified as '<50% s-WPS chance' who did not achieve s-WPS. AUC is a measure of discriminatory ability which describes the probability that a worker achieving s-WPS will be classified as having a better chance than a worker not achieving s-WPS. To ensure comparability between the models, estimates were presented both for the number of individuals with complete data and for the number of individuals with complete data on the adjacent model (e.g. in Model 3: estimates were presented for the...
161 individuals with complete data on demographic, patient-reported and clinical domains and for the 93 individuals with complete data on the MRI domain (Model 4). For each added domain of information, sensitivity, specificity, PPV and NPV were compared between the models. To further appraise the properties of the models, the chance of s-WPS was divided into four categories (< 30%, ≥ 30% and < 50%, ≥ 50% and < 70%, ≥ 70%) and tabulations were made to assess whether adding domains of information improved prediction for those with the lowest (< 30%) and highest (≥ 70%) chance of s-WPS. Estimates were reported using 95% confidence intervals (CIs).

The statistical analyses were performed by the first author who was not involved in the clinical assessment or interventions given in the original RCT [11]. STATA15 statistical software package was used for all statistical analyses (STATA Corp., College Station, Texas, USA).

Nomenclature

This study explored the associations between prognostic variables and s-WPS. The term "prognostic variable" denotes a baseline variable which is associated with the outcome s-WPS but without making inferences about causality [26]. The terms 'predict', 'prediction' relate to the ability of the models to forecast the prognosis regarding work participation.

Ethics

All participants provided written informed consent. The study was approved by The Danish Data Protection Agency (J. no. 1-16-02-86-16) and by The Central Denmark Region Committees on Health Research Ethics (M-20090027).

Results

One hundred-and-sixty-eight sickness absentees were included. S-WPS was achieved by 80 (47.6%) of the sickness absentees, i.e. 47.6% of the sickness absentees had a work participation score ≥ 75% from Week 30 to 104 after enrolment. The proportion achieving s-WPS was similar (47.4%) in the MRI sample. Baseline characteristics are presented in Table 1, and were similar in the MRI sample compared with the entire cohort (data not shown).

The crude associations between potential prognostic variables and s-WPS are presented in Table 2. For those variables yielding $p < 0.2$, the following associations were seen: Increased odds of s-WPS were found for male sex, radiculopathy and kyphosis. Decreased odds of s-WPS were found for increased sick leave duration (months),
increased ÖMPSQ score, ≥ 11 tender points, and spinal canal stenosis. The remaining variables from Table 2 did not fulfill the criterion of \( p < 0.2 \) for inclusion in the multivariable analyses.

Table 3 presents the predictive values from Models 1-4. As shown, knowledge about only the sex offered an NPV of 0.60 (0.46; 0.74) and a PPV of 0.58 (0.49; 0.67). By adding patient-reported variables (duration of sick leave and ÖMPSQ score), the NPV increased to 0.70 (0.59; 0.80) and the PPV to 0.72 (0.62; 0.81) along with increased specificity. Adding clinical variables elicited only minor changes in NPV and sensitivity while the AUC changed in the second decimal. Further addition of MRI variables did not improve any of the values.

Additional detail regarding Models 1-4 is offered by Table 4 where the chance of s-WPS in four categories is compared with actual WPS outcomes. Model 1 did not identify anybody with < 30% or ≥ 70% of s-WPS, whereas Model 2 performed better than Model 1, discriminating among those whose chance of s-WPS was very low or very high. The discriminating properties of the models were similar for Models 2, 3 and 4 (assessed by the distribution of chance).

**Discussion**

To our knowledge, this is the first study exploring the contribution of demographics, patient-reported, clinical and MRI information to the prediction of work participation in sickness absentees with neck or shoulder pain. The results suggest that demographics and patient-reported measures include important prognostic information (sex, duration of sick leave prior to enrollment and ÖMPSQ score) and that neither clinical nor MRI information provide substantial additional information to the prediction of work participation within 2 years. That is, Model 2 performed similarly to Models 3 and 4 regarding sensitivity, specificity, NPV, PPV and AUC. Model 2 also performed similarly to Models 3 and 4 at discriminating between those having the lowest and highest chance of s-WPS (Table 4).

**Comparison with other studies:**

Only 47.6% of the sickness absentees achieved s-WPS within two-year follow up, which is indicative of a population with a high level of work disability. In a large study of neck and back pain [27], only 8% reported sick leave the previous year when asked at 3-year follow up. The study [27] was conducted in a workplace setting, hence, study participants were presumably less disabled than those referred to secondary care settings as in the current study. From secondary care settings however, previous studies [6, 28] have also found what seemed to be better
work outcomes than in the current study. In a study of low back pain, 72% achieved RTW within 1 year, however RTW was defined as 4 consecutive weeks of work [6]. Another study of primarily musculoskeletal pain used a more sustainable RTW outcome, namely 3 consecutive months with increased working time compared to baseline [28]. Sixty percent had achieved this outcome at 1 year follow up and percent-wise more comparable with our study findings. Differences in population characteristics, setting or the outcome measures chosen may explain the wide range of work participation following musculoskeletal pain. Biering et al. suggested that for prognostic purposes the WPS was superior to time-to-event-based measures and measures at fixed time-points [14]. As we used the WPS in the current study and included the study population from a secondary care setting, we believe that the fact that less than half of the study participants achieved s-WPS is indicative of work disability that hampers work participation.

A comprehensive literature search strategy adapted from a 2017 review [2] was performed using Medline and yet, no studies were identified which categorized prognostic variables and used an analytical approach resembling the order of information obtained in a clinical setting. For this reason, direct comparison with previous studies was not possible, but some studies reported measures that can be compared with the current study [29, 30]. In a Norwegian study [29], the sickness absentees’ own prediction of sick leave duration ≥ 26 weeks yielded a slightly better PPV (0.78) than any of the models in the current study. However, sensitivity based on self-prediction was 0.28 and sensitivity based on dedicated medical consultants was 0.07. Both measures are notably lower than the sensitivity estimates of the current study suggesting that 1) perhaps the sickness absentees in Fleten et al.’s study were over-optimistic, leading to low sensitivity of self-prediction, and 2) the information that was available to the medical consultants [29] was inadequate to result in a sensitivity as high as in the current study.

Rehabilitation professionals in another study [30] predicted the chance of RTW based on sick leave duration, reason for sick leave, unemployment, age > 45 years, female sex and ‘gut feeling’. Their prediction was concordant with actual RTW in 73% and thus similar to the specificity estimates of the current study. Of interest is that the rehabilitation professionals in that study [30] had access to some similar information as was available in the current study, namely sex and sick leave duration. Their access to further information about reason for sick leave, unemployment, age > 45 and "gut feeling" did not yield a higher specificity than in the current study.

Regarding the association between sex and work outcomes, results in previous studies have been conflicting. In a population-based study [31] and in studies from secondary care on spinal pain [32] and shoulder pain [33], sex was not associated with work outcomes, while a primary care study [24] did find predictive value of sex. In summary,
reviews have found conflicting evidence on the impact of sex on work outcomes [1, 3] and with this in mind, the current study's estimates of PPV and NPV when based on sex alone are not surprising.

The impact of sick leave duration corroborates the findings in numerous reviews [1, 3, 7, 34, 35] and cohort studies of shoulder pain, spinal pain and whiplash trauma [24, 36, 37]. Due to the non-modifiable nature of sick leave duration, our findings highlight the need for action that prevents and/or addresses sick leave early enough to minimize the duration.

The negative association between the ÖMPSQ score and s-WPS is in line with previous findings [18, 38]. While studies from primary care settings and workplaces [39, 40] have suggested ÖMPSQ cut-off values of 90 and 105, the current study suggests that in a secondary care setting, ÖMPSQ has important properties regarding prognosis of work participation also as a numerical variable.

Interestingly, a study of primarily chronic spinal pain suggested cut-offs of 90 and 105 [40] but did not yield sensitivity and specificity that were concurrently as high as any of the models including ÖMPSQ in the current study. In contrast, a primary care study of patients with acute or sub-acute spinal pain [39], a cut-off of 90 yielded better sensitivity and slightly lower specificity [39], thereby demonstrating the properties of ÖMPSQ in early detection of poor prognosis.

In the context of clinical variables, a cohort study on low back pain [41] found a crude association between tender points count (numeric variable) and unsuccessful RTW. However, this was not maintained in the multivariable model. Moreover, no association was found between radiculopathy and work outcomes [41], which is supported by a review presenting moderate evidence for no association between radiating pain and RTW [2]. Therefore, our finding of the limited improvement of predictive values in Model 3 is in line with these previous findings [2, 41].

Finally, regarding MRI, only one study was identified which explored the association between MRI findings and work outcomes [6]. In this study of sickness absentees with low back pain, the presence of Modic type 1 changes was associated with unsuccessful RTW. Unfortunately, we were unable to assess the impact of Modic changes in the current study since the inter- and intra-rater reliability for this pathology was not established owing to too low prevalences (unpublished).

**Methodological considerations**

WPS was chosen as the outcome measure in an attempt to capture both work reintegration and work maintenance although we are aware that other RTW outcomes reflect other aspects regarding the process and context of RTW
The chosen cut-off value (s-WPS ≥ 75% or not) showed moderate to substantial agreement with other RTW measures (RTW yes/no at a given time point and time to 4 weeks of self-support without relapses, respectively) in a previous study [14]. As previously argued, the same study [14] found WPS suitable for prognostic purposes. When classifying the chance of s-WPS, a 50% threshold was chosen for two reasons: 1) A reasonable balance between sensitivity and specificity measures was desired and 2) Our clinical experience suggests that sickness absentees with neck or shoulder pain who are seen in secondary care are troubled with many biopsychosocial issues. Hence, choosing higher cut-offs for the chance of s-WPS might be too optimistic. A different cut-off would probably be relevant if the study population had comprised patients from primary care on short-term sick leave.

**Strengths**

First, the use of registry data for the outcome assessment ensured 100% follow up thereby reducing the risk of attrition bias. Second, the use of registry data reduced the risk of measurement bias since the outcome assessment was unaffected by knowledge of the prognostic variables. Likewise, the assessment of prognostic factors was unaffected by knowledge of the outcome. Third, the risk of attrition bias was minimized by the low number of missing values for the majority of variables (demographic, patient-reported and clinical variables). Only for MRI was the number of missing values substantial. However, the distribution between exposure and outcome was assessed for all the variables in Table 1, which revealed that attrition was not skewed for any of the variables, hence attrition bias was not suspected (data not shown). Fourth, the duration of follow-up reaching 2 years constitutes a realistic long-term outcome. A fifth strength is the a priori decision to take an analytical approach resembling the working conditions of the clinicians who are responsible for appraisal of work prognosis. This reduced the risk of purely data-driven results. Finally, when bearing in mind the well-established impact of external societal factors on the process of sick leave [43, 44], it is a further strength of the current study that the Danish legislation on sickness benefits did not undergo major changes from 2009-2014. A change in legislation in 2014 reduced the right for sickness benefits from 52 to 22 weeks. It affected only six sickness absentees (equally distributed between s-WPS and u-WPS) and is thus not suspected to bias the estimates.

**Limitations**

First, since the study population was originally included in an RCT, the results may not be generalizable to all sickness absentees with neck or shoulder pain. The possibility of sampling bias due to referral patterns of the general
practitioners cannot be refuted nor elucidated. Hence, the results are expected to be representative of sickness
absentees with neck or shoulder pain who are seen in a secondary care setting. Generalizability to primary care
settings should be made with caution and confirmatory studies including all sickness absentees with neck or
shoulder pain would be needed to improve generalizability.

Second, the representativeness of the results should be considered in the light of the above-mentioned change in
legislation, implying that the right to sickness benefits is now limited to 22 weeks compared with 52 weeks during
most of the study. Confirmatory studies would be desirable to assess the impact of this change in legislation. But
given that the literature over the past decades is corroborated (i.e. major impact of sick leave duration and
psychosocial factors [45]), we expect that similar results would be found.

Third, the limited MRI sample affects precision of the estimates; an issue that was further attenuated for certain MRI
variables (disc bulge/protrusion/extrusion, neural foraminal stenosis, zygapophyseal osteoarthritis and uncovertebral
osteoarthritis). For these variables, the available sequences did not allow for evaluation on all 97 MRIs (Table 1 and
(unpublished). Had MRI been available for the entire cohort, confidence intervals for MRI variables would have
been narrowed down.

Fourth, the use of registry data is usually an advantage, but the data source also warrants consideration since
registration procedures imply a risk of unequal registration of short-term sick leave. Sick leave registration in
DREAM begins at the end of the employer paid period and backward adjustments are made, so that the number of
sickness benefit weeks in DREAM equal the total number of sick leave weeks. Since registration is initiated at the
end of the employer paid period, multiple absences due to sick leave lasting only days or a few weeks are usually
not registered (because they are within the employer-paid period). In the current study, some participants may have
been misclassified as having achieved s-WPS although they had multiple short-term absences due to sick leave. A
related problem applies to employees with a §56-agreement, i.e. their employers are entitled to reimbursement of
sickness benefits from day one; such employees may be misclassified as u-WPS in the current study due to sick
leave registrations exceeding actual duration. These possible misclassifications are considered non-differentiated
since they are not suspected to be associated with the exposure (i.e. the prognostic variables). Unfortunately, the
data do not offer any insight as to the possible distribution of §56-agreements in the current study. However,
explorative post-hoc analyses were performed in which single weeks of sickness benefit reimbursement were
considered as §56-agreements, i.e. regarded as working weeks. This resulted in the distribution of s-WPS/u-WPS
changing from 80/88 to 83/85 (data not shown). It did not change the results of the study.
Sixth, the limited sample size implies a risk of type 2 errors. That is, important prognostic factors may not be discovered as statistically significant owing to the limited sample size. Other factors that could affect the prognosis for work outcomes include fear avoidance beliefs [2, 3, 45] and physical workload [1, 2, 35]. These factors were not isolated for analysis in the current study since they were covered by the ÖMPSQ score and we wished for all variables to be as mutually exclusive as possible. Furthermore, the aim was not to investigate a complete list of all possible factors affecting WPS prognosis but rather to explore the contributions of demographic, patient-reported, clinical and MRI variables for which purpose we believe the current sample size was adequate.

Conclusion

In sickness absentees with neck or shoulder pain, clinical and MRI information provide no additional information for the prediction of work participation compared with only demographic and patient-reported information. Though Model 2 performs similarly to Models 3 and 4, there are limitations with respect to discriminating between those with the lowest and highest chance of s-WPS (Table 4). This means that clinical applicability requires caution if the aim is identification of those who have the lowest/highest chances of s-WPS. The results do, however, provide valuable knowledge to clinicians both in the assessment of work prognosis and in dialogue with patients and other stakeholders: prediction should primarily be based on demographic and patient-reported information, not on clinical and MRI findings.

Due to the exploratory nature of the current study, the generalizability of findings needs to be assessed in independent studies, preferably in both primary and secondary care settings. Finally, from a public health perspective, there is a need for management of neck and shoulder pain and identification of those at risk of poor work prognosis early enough to avoid harmful long-term absences due to sick leave.

All procedures followed were in accordance with the ethical standards of The Central Denmark Region Committees on Health Research Ethics. Informed consent was obtained from all participants included in the study.

Abbreviations

AUC: Area Under the Curve, MRI: Magnetic Resonance Imaging, NPV: Negative Predictive Value, ÖMPSQ: Örebro Musculoskeletal Pain Screening Questionnaire, PPV: Positive Predictive Value, RCT: Randomized
Controlled Trial, RTW: Return To Work, s-WPS: successful Work Participation Score, u-WPS: unsuccessful Work Participation Score.

Tables
Table 1 Baseline characteristics in sickness absentees with neck or shoulder pain.
Table 2 Odds ratios for work participation score ≥ 75% (s-WPS) in sickness absentees with neck or shoulder pain.
Table 3 Predictive values, sensitivity, specificity and AUC for the prediction of work participation score < 75% (u-WPS) or ≥ 75% (s-WPS) in 4 models reflecting type of prognostic information.
Table 4 Cross-tabulations of the distribution of chance vs. achieved work participation score ≥ 75% (s-WPS) or < 75% (u-WPS).

Authors' contributions
Line Thorndal Moll: Design of the study, analysis and interpretation of data, primary responsible for the manuscript.
Anne Mette Schmidt: Interpretation of data, critical appraisal of the manuscript.
Christina Malmose Stapelfeldt: Design of the study, interpretation of data, critical appraisal of the manuscript.
Merete Labriola: Design of the study, interpretation of data, critical appraisal of the manuscript.
Ole Kudsk Jensen: Design of the study, interpretation of data, critical appraisal of the manuscript.
Morten Wasmod Kindt: Interpretation of data, critical appraisal of the manuscript.
Tue Secher Jensen: Design of the study, interpretation of data, critical appraisal of the manuscript.
Berit Schiöttz-Christensen: Design of the study, interpretation of data, critical appraisal of the manuscript.

Conflicts of interest
Author Line Thorndal Moll declares that she has no conflict of interest.
Author Anne Mette Schmidt declares that she has no conflict of interest.
Author Christina Malmose Stapelfeldt declares that she has no conflict of interest.
Author Merete Labriola declares that she has no conflict of interest.
Author Ole Kudsk Jensen declares that he has no conflict of interest.
Author Morten Wasmod Kindt declares that he has no conflict of interest.
Author Tue Secher Jensen declares that he has no conflict of interest.
Author Berit Schiöttz-Christensen declares that she has no conflict of interest.
Table 1 Baseline characteristics in sickness absentees with neck or shoulder pain.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>N</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex, n (%)</td>
<td>168</td>
<td>53 (31.5)</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>168</td>
<td>40.5 (34.5 ; 48.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient-reported information</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain intensity (0-10) last week, median (IQR)</td>
<td>158</td>
<td>7 (5 ; 8)</td>
</tr>
<tr>
<td>Whiplash trauma, n (%)</td>
<td>168</td>
<td>24 (14.3)</td>
</tr>
<tr>
<td>Sick leave duration (months), median (IQR)</td>
<td>168</td>
<td>2.3 (1.2 ; 3.0)</td>
</tr>
<tr>
<td>ÖMPSQ score, mean (SD)</td>
<td>161</td>
<td>122 (23.7)</td>
</tr>
<tr>
<td>Education ≥ 3 years, n (%)</td>
<td>155</td>
<td>29 (18.7)</td>
</tr>
<tr>
<td>Ongoing worker's compensation claim, n (%)</td>
<td>139</td>
<td>36 (25.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical information</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiculopathy, n (%)</td>
<td>168</td>
<td>41 (24.4)</td>
</tr>
<tr>
<td>≥ 11 tender points, n (%)</td>
<td>168</td>
<td>57 (33.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intervention group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief intervention, n (%)</td>
<td>168</td>
<td>83 (49.4)</td>
</tr>
<tr>
<td>Multidisciplinary intervention, n (%)</td>
<td>168</td>
<td>85 (50.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MRI information</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis, n (%)</td>
<td>97</td>
<td>23 (23.7)</td>
</tr>
<tr>
<td>Disc height reduction, n (%)</td>
<td>97</td>
<td>71 (73.2)</td>
</tr>
<tr>
<td>Bulge, protrusion or extrusion, n (%)</td>
<td>83</td>
<td>69 (83.1)</td>
</tr>
<tr>
<td>Spinal canal stenosis, n (%)</td>
<td>97</td>
<td>14 (14.4)</td>
</tr>
<tr>
<td>Neural foraminal stenosis, n (%)</td>
<td>84</td>
<td>46 (54.8)</td>
</tr>
<tr>
<td>Zygapophyscal osteoarthritis, n (%)</td>
<td>83</td>
<td>39 (47.0)</td>
</tr>
<tr>
<td>Uncovertebral osteoarthritis, n (%)</td>
<td>78</td>
<td>37 (47.4)</td>
</tr>
</tbody>
</table>
Table 2. Odds ratios for work participation score ≥ 75% (s-WPS) in sickness absentee in those with neck or shoulder pain.

<table>
<thead>
<tr>
<th>Demographic information</th>
<th>N</th>
<th>Odds ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>168</td>
<td>2.13 (1.10; 4.13)*</td>
<td>0.03</td>
</tr>
<tr>
<td>Age (years)</td>
<td>168</td>
<td>0.99 (0.96; 1.03)</td>
<td>0.75</td>
</tr>
<tr>
<td>Patient-reported information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain intensity (numeric rating scale)</td>
<td>161</td>
<td>0.99 (0.96, 1.02)</td>
<td>0.38</td>
</tr>
<tr>
<td>Whiplash trauma (yes vs. no)</td>
<td>168</td>
<td>0.76 (0.51; 1.18)</td>
<td>0.53</td>
</tr>
<tr>
<td>Sick leave duration (months)</td>
<td>168</td>
<td>0.60 (0.46; 0.78)*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>OMPSQ score (numerical variable)</td>
<td>161</td>
<td>0.97 (0.95; 0.98)*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Education ≥ 3 years (yes vs. no)</td>
<td>155</td>
<td>1.22 (0.54; 2.73)</td>
<td>0.63</td>
</tr>
<tr>
<td>Ongoing worker’s compensation claim</td>
<td>139</td>
<td>0.65 (0.30; 1.40)</td>
<td>0.27</td>
</tr>
<tr>
<td>Clinical information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiculopathy (yes vs. no)</td>
<td>168</td>
<td>2.54 (1.22; 5.33)*</td>
<td>0.01</td>
</tr>
<tr>
<td>≥ 11 tender points (yes vs. no)</td>
<td>168</td>
<td>0.51 (0.27; 0.99)*</td>
<td>0.05</td>
</tr>
<tr>
<td>Multidisciplinary intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyphosis (yes vs. no)</td>
<td>97</td>
<td>2.6 (0.98; 6.89)*</td>
<td>0.06</td>
</tr>
<tr>
<td>Disc height reduction (yes vs. no)</td>
<td>97</td>
<td>1.07 (0.44; 2.64)</td>
<td>0.88</td>
</tr>
<tr>
<td>Disc contour change (bulge, protrusion or extrusion) (yes vs. no)</td>
<td>83</td>
<td>1.30 (0.40; 4.13)</td>
<td>0.66</td>
</tr>
<tr>
<td>Spinal canal stenosis (yes vs. no)</td>
<td>97</td>
<td>0.39 (0.11; 1.33)*</td>
<td>0.14</td>
</tr>
<tr>
<td>Neural foraminal stenosis (yes vs. no)</td>
<td>84</td>
<td>1.13 (0.48; 2.68)</td>
<td>0.78</td>
</tr>
<tr>
<td>Zygapophysial osteoarthritis (yes vs. no)</td>
<td>83</td>
<td>0.76 (0.32; 1.82)</td>
<td>0.54</td>
</tr>
<tr>
<td>Uncocervical osteoarthritis (yes vs. no)</td>
<td>78</td>
<td>1.21 (0.50; 2.95)</td>
<td>0.68</td>
</tr>
</tbody>
</table>

* indicate variables carried forward in the multivariable analyses.
OMPSQ: Örebro Musculoskeletal Pain Screening Questionnaire
MRI: Magnetic Resonance Imaging
Table 3. Predictive values, sensitivity, specificity and AUC for the prediction of work participation score < 75% (u-WPS) or ≥ 75% (s-WPS) in 4 models reflecting type of prognostic information

<table>
<thead>
<tr>
<th>Information included</th>
<th>N</th>
<th>Classified as ≥ 50% chance and achieved s-WPS (NPV)</th>
<th>Classified as &lt; 50% chance and achieved u-WPS (PPV)</th>
<th>s-WPS correctly classified (specificity)</th>
<th>u-WPS correctly classified (sensitivity)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Demographic*</td>
<td>168</td>
<td>0.50 [ (0.46 ; 0.74) ]</td>
<td>0.58 [ (0.49 ; 0.67) ]</td>
<td>0.40 [ (0.29 ; 0.52) ]</td>
<td>0.76 [ (0.66 ; 0.85) ]</td>
<td>0.58 [ (0.51 ; 0.65) ]</td>
</tr>
<tr>
<td>Model 1: Demographic</td>
<td>161</td>
<td>0.51 [ (0.46 ; 0.74) ]</td>
<td>0.59 [ (0.49 ; 0.68) ]</td>
<td>0.41</td>
<td>0.76</td>
<td>0.59</td>
</tr>
<tr>
<td>Model 2: Demographic</td>
<td>161</td>
<td>0.70 [ (0.49 ; 0.83) ]</td>
<td>0.72 [ (0.62 ; 0.81) ]</td>
<td>0.68 [ (0.50 ; 0.83) ]</td>
<td>0.74 [ (0.65 ; 0.83) ]</td>
<td>0.76</td>
</tr>
<tr>
<td>Model 3: Demographic + patient-reported**</td>
<td>161</td>
<td>0.69 [ (0.62 ; 0.84) ]</td>
<td>0.73 [ (0.63 ; 0.82) ]</td>
<td>0.57 [ (0.57 ; 0.79) ]</td>
<td>0.69 [ (0.65 ; 0.83) ]</td>
<td>0.68 [ (0.68 ; 0.83) ]</td>
</tr>
<tr>
<td>Model 3: Demographic + patient-reported + clinical***</td>
<td>161</td>
<td>0.52 [ (0.50 ; 0.83) ]</td>
<td>0.82 [ (0.63 ; 0.82) ]</td>
<td>0.65 [ (0.55 ; 0.77) ]</td>
<td>0.69 [ (0.65 ; 0.84) ]</td>
<td>0.77</td>
</tr>
<tr>
<td>Model 4: Demographic + patient-reported + clinical + MRI****</td>
<td>93</td>
<td>0.85 [ (0.51 ; 0.80) ]</td>
<td>0.73 [ (0.58 ; 0.82) ]</td>
<td>0.70 [ (0.54 ; 0.83) ]</td>
<td>0.72 [ (0.53 ; 0.82) ]</td>
<td>0.77</td>
</tr>
</tbody>
</table>

*: male sex OR = 2.1 (95% CI: 1.1 ; 4.1).
**: male sex OR = 2.0 (95% CI: 0.96 ; 4.2), sick leave (months) OR = 0.63 (95% CI: 0.49 ; 0.86), ÖMPSQ score OR = 0.97 (95% CI: 0.96 ; 0.99).
***: male sex OR = 1.6 (95% CI: 0.74 ; 3.6), sick leave (months) OR = 0.65 (95% CI: 0.49 ; 0.87), ÖMPSQ score OR = 0.97 (95% CI: 0.96 ; 0.99), radiculopathy OR = 1.74 (95% CI: 0.73 ; 4.1), ≥ 11 tender points OR = 0.8 (95% CI: 0.38 ; 1.7).
****: male sex OR = 2.8 (95% CI: 0.88 ; 8.9), sick leave (months) OR = 0.53 (95% CI: 0.34 ; 0.85), ÖMPSQ score OR = 0.98 (95% CI: 0.96 ; 1.0), radiculopathy OR = 1.54 (95% CI: 0.42 ; 5.71), ≥ 11 tender points OR = 0.8 (95% CI: 0.30 ; 2.2), kyphosis OR = 1.6 (95% CI: 0.45 ; 5.8), spinal canal stenosis OR = 0.2 (95% CI: 0.03 ; 1.1).

NPV: negative predictive value
PPV: positive predictive value
AUC: area under the curve

2. The reduction of N in all models is explained by the number of individuals with a available ÖMPSQ score for the given models.
<table>
<thead>
<tr>
<th>Domains included</th>
<th>N*</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Chance of s-WPS</td>
</tr>
<tr>
<td>Demographics*</td>
<td>168</td>
<td>&lt; 30%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30% to 49%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50% to 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 70%</td>
</tr>
<tr>
<td>Demographic + patient-reported**</td>
<td>161</td>
<td>&lt; 30%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30% to 49%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50% to 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 70%</td>
</tr>
<tr>
<td>Demographic + patient-reported + clinical***</td>
<td>161</td>
<td>&lt; 30%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30% to 49%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50% to 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 70%</td>
</tr>
<tr>
<td>Demographic + patient-reported + clinical + MRI****</td>
<td>93</td>
<td>&lt; 30%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30% to 49%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50% to 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 70%</td>
</tr>
</tbody>
</table>

*: The reduction of N in all models is explained by the number of individuals with available OMPSQ score for the given models.
References


22


Evaluation manual for the inter- and intra-rater reliability study on MRI of the cervical spine

This evaluation manual covers all the definitions and classifications that are used for the above mentioned reliability study.
In this evaluation manual, the names of the variables are so-called 'generic' names, whereas variables in Epidata have appropriate prefixes to determine the cervical level in question. Likewise, for the neural foramina, the uncovertebral and zygapophyseal joints, suffixes in Epidata are used to denote whether it is on the left or the right hand side.

Six disc levels are examined from C2/C3 to C7/T1.

Kyphosis

0 : normal or reduced lordosis
1 : kyphosis
9 : missing

Cervical alignment is assessed on mid-sagittal T2-weighted images. A line from the posteroinferior aspect of vertebra C2 to the posteroinferior aspect of vertebra C7 is drawn. If any part of C3-C6 lie behind/posterior to this line, the alignment is classified as kyphosis (1).

Disc height

0 : normal disc height
1 : reduced disc height
9 : missing

Disc height is assessed on T2-weighted sagittal images.
This classification is inspired by Jacobs et al. (2) but modified since the relatively small anatomical structures in the cervical spine complicate assessments of nucleus signal.

Reduced height is assessed by comparison with neighbouring discs that look morphologically normal (3). In case of extensive disc degeneration thus complicating between-disc comparison, the assessment is based on the experience of the reader.

Both of these have normal disc height = grade 0. From Jacobs et al. (2).
Reduced height = grade 1

Reduced height = grade 1
Disc contour

0 : normal
1 : bulge or protrusion
2 : extrusion
9 : missing

Disc contour is assessed by use of both sagittal and axial images. We use a modified version of the classification by Fardon et al. 2014 (4).

- A normal disc is defined as morphologically normal. I.e., there is no sign of illness, trauma or ageing. The disc is located within the "disc space", the limits of which are determined by the vertebral endplates craniocaudally. The peripheral limits of "disc space" are constituted by the apophyses of the vertebrae (excluding possible osteophytes) (4).

- Disc bulge describes a general (in some cases asymmetric) bulging of disc material. The material has moved beyond "disc space" and comprises > 25% of the disc circumference. Usually, a bulge does not exceed 3 mm beyond "disc space" (4).

- Protrusion: A focal displacement of disc material of which the base involves < 25% of the circumference. The maximum measure of the displaced disc material is smaller than the measure of the base of the displaced material at the disc space of origin, measured in the same plane (4).

- Extrusion: A focal displacement of disc material where the base involves < 25% of the disc circumference. The maximal measure of the displaced disc material is greater than the measure of the base of the displaced material at the disc space of origin, measured in the same plane. Extrusion also comprises what was formerly denoted as sequestration where the displaced material has lost continuity with the disc space of origin (4).
Normal discs on axial and sagittal images

Fig. 1. Normal lumbar disc. (Top Left) Axial, (Top Right) sagittal, and (Bottom) coronal images demonstrate that the normal disc, composed of central NP and peripheral AF, is wholly within the boundaries of the disc space, as defined, cranially and caudally by the vertebral body end plates and peripherally by the planes of the outer edges of the vertebral apophyses, exclusive of osteophytes. NP, nucleus pulposus; AF, annulus fibrosus.
Fig. 3. Bulging disc. (Top Left) Normal disc (for comparison); no disc material extends beyond the periphery of the disc space, depicted here by the broken line. (Top Right) Symmetric bulging disc; annular tissue extends usually by less than 1 mm, beyond the edges of the vertebral apophysis symmetrically throughout the circumference of the disc. (Bottom) Asymmetric bulging disc; annular tissue extends beyond the edges of the vertebral apophysis, asymmetrically greater than 25% of the circumference of the disc.
Fig. 4. Herniated disc: protrusion. (Left) Axial and (Right) sagittal images demonstrate displaced disc material extending beyond less than 25% of the disc space, with the greatest measure, in any plane, of the displaced disc material being less than the measure of the base of displaced disc material at the disc space of origin, measured in the same plane.
Extrusion, axial

Extrusion, sagittal

Fig. 5. Herniated disc extrusion. (Left) Axial and (Right) sagittal images demonstrate that the greatest measure of the displaced disc material is greater than the base of the displaced disc material at the disc space of origin, when measured in the same plane.
**Spinal canal stenosis**

0: no stenosis  
1: >50% obliteration of CSF, no cord deformity  
2: >50% obliteration of CSF with cord deformity but no signal change  
3: >50% obliteration of CSF with cord deformity and signal change  
9: missing

The spinal canal is assessed on three T2-weighted images; the mid-sagittal and the two neighbouring slices (one on each side of the mid-sagittal image). The classification by Kang et al. is used (5).
Vertebral endplate signal change (VESC)

0 : normal
1 : VESC type 1
2 : VESC type 2
3 : VESC type 3
4 : mixed VESC type 1 and 2
9 : missing

Primarily assessed on sagittal images. VESC constitute changes in the endplates of two neighbouring vertebral endplates surrounding a degenerated disc (4). If only visible on one slice, the finding is excluded (6). Likewise, changes only related to osteophytes or Schmorl nodes will be excluded.

Type 1 is hyperintense on T2-weighted and hypo-/isointense on T1-weighted images (7).
Type 2 is hyperintense on T1-weighted images and hyper-/isointense on T2-weighted images (7).
Type 3 is hypointense on T1 and T2 (bone sclerosis where hydrogen protons are so tightly bound and thus difficult to bring in motion → no MR signal on T1 and T2) (4).
VESC type 1. T2-weighted

VESC type 2. T1-weighted

VESC type 2. T2-weighted
Neural foraminal stenosis

0 : normal or < 50% obliteration of perineural fat
1 : >= 50% obliteration of perineural fat with/without nerve root compromise.
9 : missing

The extent of perineural fat obliteration is determined on T2 weighted oblique images. The assessment is made on the slice with most severe stenosis. Meanwhile, on the axial images, the reader ensures that he/she is actually looking at the neural foramen. The boundaries of the neural foramen are: to the anterior the disc and posterior aspect of the vertebral bodies; to the posterior the zygapophyseal joint; and cranially and caudally the pediculi (8). On axial images, the medial and lateral boundaries are the medial and lateral aspect of the pediculi, respectively (9).

If no oblique images are available, the assessment is based on the axial images where stenosis is most severe, i.e. where the distance between the processus articularis superior (posterolaterally) and corpus vertebra/discus (anteromedially) is identified (10).
In the following, foraminal stenosis on axial images is depicted. While assessing the axial images, the sagittal images are used to demonstrate the location in relation to the processes articulares. The severity of stenosis is determined by comparing the relevant number of subsequent axial slices in the neural foramen in question.

Grade 0 (15):
0 = figure a +b

Grade 1(15):
1 = figure c+d
Uncovertebral osteoarthritis

0 : normal
1 : definite osteoarthritis (primarily assessed by the presence of osteophytes which take up place in the neural foramen).
9 : missing

Only if $\geq 50\%$ foraminal stenosis is settled on, the reader can consider the classification definite osteoarthritis. (i.e. if foraminal stenosis is $<50\%$, any possible irregularities of the uncinate process are not articulate enough to be classified as osteoarthritis). The classification was chosen because degenerative changes of the uncovertebral joint primarily comprise the growth of osteophytes (11).

Assessment is only done on oblique images which allow for the best assessments (11,12). If these are not available, the value '9' = missing is used due to the risk of partial volume effect on axial images.
Zygapophyseal osteoarthritis

0: normal (no definite joint space narrowing, osteophytes or hypertrophy of the processus articularis)
1: definite osteoarthritis (definite joint space narrowing, osteophytes or hypertrophy of the processus articularis) (13,14)
9: missing

The assessment is done on oblique images (12). If these are not available, axial and sagittal images are used. **Only if** >= 50% foraminal stenosis is settled on, the reader can consider the classification definite osteoarthritis. (i.e. if foraminal stenosis is <50%, any possible irregularities of the zygapophyseal joint are not articulate enough to be classified as osteoarthritis.)
Comments
If any other relevant pathological findings are identified. This may be spondylolistesis, fracture, cord pathology etc.


### Table 5. Prevalence of positive MRI findings

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>Number of MRIs assessed</th>
<th>Reader A, 1st assessment</th>
<th>Reader B</th>
<th>Reader C</th>
<th>Number of MRIs assessed</th>
<th>Positive findings</th>
<th>Reader A, 2nd assessment</th>
<th>Positive findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyphosis</td>
<td></td>
<td>50</td>
<td>15 (30.0)</td>
<td>50</td>
<td>17 (34.0)</td>
<td>50</td>
<td>20 (40.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C2/C3</td>
<td></td>
<td>50</td>
<td>3 (6.0)</td>
<td>50</td>
<td>0 (0.0)</td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C3/C4</td>
<td></td>
<td>50</td>
<td>8 (16.0)</td>
<td>50</td>
<td>4 (8.0)</td>
<td>50</td>
<td>8 (16.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C4/C5</td>
<td></td>
<td>50</td>
<td>12 (24.0)</td>
<td>50</td>
<td>15 (30.0)</td>
<td>50</td>
<td>21 (42.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C5/C6</td>
<td></td>
<td>50</td>
<td>26 (52.0)</td>
<td>50</td>
<td>24 (48.0)</td>
<td>50</td>
<td>32 (64.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C6/C7</td>
<td></td>
<td>50</td>
<td>21 (42.0)</td>
<td>50</td>
<td>16 (32.0)</td>
<td>50</td>
<td>22 (44.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc height C7/T1</td>
<td></td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
<td>2 (4.0)</td>
<td>50</td>
</tr>
<tr>
<td>Disc contour C2/C3</td>
<td></td>
<td>30</td>
<td>1 (3.3)</td>
<td>50</td>
<td>1.20</td>
<td>50</td>
<td>1 (2.0)</td>
<td>31</td>
</tr>
<tr>
<td>Disc contour C3/C4</td>
<td></td>
<td>38</td>
<td>11 (28.9)</td>
<td>50</td>
<td>9 (18.0)</td>
<td>50</td>
<td>8 (16.0)</td>
<td>35</td>
</tr>
<tr>
<td>Disc contour C4/C5</td>
<td></td>
<td>45</td>
<td>15 (33.3)</td>
<td>50</td>
<td>13 (26.0)</td>
<td>50</td>
<td>15 (30.0)</td>
<td>45</td>
</tr>
<tr>
<td>Disc contour C5/C6</td>
<td></td>
<td>48</td>
<td>31 (64.6)</td>
<td>50</td>
<td>26 (52.0)</td>
<td>50</td>
<td>32 (64.0)</td>
<td>48</td>
</tr>
<tr>
<td>Disc contour C6/C7</td>
<td></td>
<td>46</td>
<td>26 (56.5)</td>
<td>50</td>
<td>19 (38.0)</td>
<td>50</td>
<td>24 (48.0)</td>
<td>46</td>
</tr>
<tr>
<td>Disc contour C7/T1</td>
<td></td>
<td>38</td>
<td>2 (5.2)</td>
<td>50</td>
<td>3 (6.0)</td>
<td>50</td>
<td>1 (2.0)</td>
<td>34</td>
</tr>
<tr>
<td>Spinal canal stenosis C2/C3</td>
<td>50</td>
<td>0 (0.0)</td>
<td>50</td>
<td>0 (0.0)</td>
<td>49</td>
<td>1 (2.0)</td>
<td>50</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----</td>
<td>---------</td>
<td>----</td>
<td>---------</td>
<td>----</td>
<td>---------</td>
<td>----</td>
<td>---------</td>
</tr>
<tr>
<td>Spinal canal stenosis C3/C4</td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
<td>1 (2.0)</td>
<td>49</td>
<td>1 (2.0)</td>
<td>50</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Spinal canal stenosis C4/C5</td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
<td>1 (2.0)</td>
<td>50</td>
<td>5 (10.0)</td>
<td>50</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Spinal canal stenosis C5/C6</td>
<td>50</td>
<td>7 (14.0)</td>
<td>50</td>
<td>6 (12.0)</td>
<td>50</td>
<td>8 (16.0)</td>
<td>50</td>
<td>6 (12.0)</td>
</tr>
<tr>
<td>Spinal canal stenosis C6/C7</td>
<td>50</td>
<td>7 (14.0)</td>
<td>50</td>
<td>7 (14.0)</td>
<td>50</td>
<td>8 (16.0)</td>
<td>50</td>
<td>5 (10.0)</td>
</tr>
<tr>
<td>Spinal canal stenosis C7/T1</td>
<td>50</td>
<td>0 (0.0)</td>
<td>50</td>
<td>0 (0.0)</td>
<td>50</td>
<td>0 (0.0)</td>
<td>50</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

| VESC C2/C3 | 50 | 0 (0.0) | 50 | 1 (2.0) | 50 | 0 (0.0) | 50 | 0 (0.0) |
| VESC C3/C4 | 50 | 1 (2.0) | 50 | 1 (2.0) | 50 | 1 (2.0) | 50 | 1 (2.0) |
| VESC C4/C5 | 50 | 0 (0.0) | 50 | 1 (2.0) | 50 | 1 (2.0) | 50 | 0 (0.0) |
| VESC C5/C6 | 50 | 2 (4.0) | 50 | 2 (4.0) | 50 | 3 (6.0) | 50 | 3 (6.0) |
| VESC C6/C7 | 50 | 3 (6.0) | 50 | 4 (8.0) | 50 | 5 (10.0)| 50 | 5 (10.0)|
| VESC C7/T1 | 50 | 0 (0.0) | 50 | 0 (0.0) | 50 | 0 (0.0) | 50 | 0 (0.0) |

VESC: vertebral endplate signal change (Modic change)
### Table 5. Prevalence of positive MRI findings continued

<table>
<thead>
<tr>
<th>MRI finding</th>
<th>Reader A, 1st assessment</th>
<th>Reader B</th>
<th>Reader C</th>
<th>Reader A, 2nd assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of MRIs assessed</td>
<td>Positive findings</td>
<td>Number of MRIs assessed</td>
<td>Positive findings</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C2/C3</td>
<td>44</td>
<td>1 (2.3)</td>
<td>43</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C3/C4</td>
<td>47</td>
<td>1 (2.1)</td>
<td>44</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C4/C5</td>
<td>47</td>
<td>9 (19.1)</td>
<td>46</td>
<td>5 (10.9)</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C5/C6</td>
<td>47</td>
<td>12 (25.5)</td>
<td>46</td>
<td>8 (17.4)</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C6/C7</td>
<td>48</td>
<td>11 (22.9)</td>
<td>46</td>
<td>10 (21.7)</td>
</tr>
<tr>
<td>Right uncovertebral osteoarthritis C7/T1</td>
<td>45</td>
<td>0 (0.0)</td>
<td>46</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C2/C3</td>
<td>44</td>
<td>0 (0.0)</td>
<td>44</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C3/C4</td>
<td>47</td>
<td>3 (6.4)</td>
<td>44</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C4/C5</td>
<td>48</td>
<td>9 (18.8)</td>
<td>46</td>
<td>4 (8.7)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C5/C6</td>
<td>47</td>
<td>11 (23.4)</td>
<td>46</td>
<td>9 (19.6)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C6/C7</td>
<td>48</td>
<td>8 (16.7)</td>
<td>46</td>
<td>5 (10.9)</td>
</tr>
<tr>
<td>Left uncovertebral osteoarthritis C7/T1</td>
<td>45</td>
<td>0 (0.0)</td>
<td>46</td>
<td>1 (2.2)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C2/C3</td>
<td>46</td>
<td>1 (2.2)</td>
<td>43</td>
<td>1 (2.3)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C3/C4</td>
<td>48</td>
<td>0 (0.0)</td>
<td>44</td>
<td>3 (6.8)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C4/C5</td>
<td>48</td>
<td>6 (12.5)</td>
<td>46</td>
<td>6 (13.0)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C5/C6</td>
<td>48</td>
<td>7 (14.6)</td>
<td>46</td>
<td>6 (13.0)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C6/C7</td>
<td>48</td>
<td>10 (20.8)</td>
<td>46</td>
<td>7 (15.2)</td>
</tr>
<tr>
<td>Right zygapophyseal osteoarthritis C7/T1</td>
<td>47</td>
<td>0 (0.0)</td>
<td>46</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C2/C3</td>
<td>46</td>
<td>0 (0.0)</td>
<td>44</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C3/C4</td>
<td>48</td>
<td>0 (0.0)</td>
<td>44</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C4/C5</td>
<td>48</td>
<td>5 (10.4)</td>
<td>46</td>
<td>6 (13.0)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C5/C6</td>
<td>48</td>
<td>9 (18.8)</td>
<td>46</td>
<td>7 (15.2)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C6/C7</td>
<td>48</td>
<td>8 (16.7)</td>
<td>46</td>
<td>5 (10.9)</td>
</tr>
<tr>
<td>Left zygapophyseal osteoarthritis C7/T1</td>
<td>47</td>
<td>0 (0.0)</td>
<td>46</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Right neural foraminal stenosis C2/C3</td>
<td>42</td>
<td>1 (2.4)</td>
<td>42</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Right neural foraminal stenosis C3/C4</td>
<td>46</td>
<td>1 (2.2)</td>
<td>44</td>
<td>1 (2.3)</td>
</tr>
</tbody>
</table>

Note: Results are presented as number of MRIs assessed (n) and prevalence (%).
<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right neural foraminal stenosis C4/C5</td>
<td>48</td>
<td>10 (20.8)</td>
<td>45</td>
<td>8 (17.8)</td>
<td>49</td>
<td>8 (16.3)</td>
<td>48</td>
<td>6 (12.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right neural foraminal stenosis C5/C6</td>
<td>48</td>
<td>13 (27.1)</td>
<td>46</td>
<td>10 (21.7)</td>
<td>49</td>
<td>14 (28.6)</td>
<td>48</td>
<td>15 (31.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right neural foraminal stenosis C6/C7</td>
<td>48</td>
<td>15 (31.3)</td>
<td>46</td>
<td>11 (23.9)</td>
<td>49</td>
<td>11 (22.4)</td>
<td>48</td>
<td>15 (31.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right neural foraminal stenosis C7/T1</td>
<td>45</td>
<td>0 (0.0)</td>
<td>46</td>
<td>0 (0.0)</td>
<td>48</td>
<td>2 (4.2)</td>
<td>45</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C2/C3</td>
<td>42</td>
<td>1 (2.4)</td>
<td>44</td>
<td>0 (0.0)</td>
<td>47</td>
<td>0 (0.0)</td>
<td>43</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C3/C4</td>
<td>46</td>
<td>3 (6.5)</td>
<td>44</td>
<td>0 (0.0)</td>
<td>49</td>
<td>4 (8.2)</td>
<td>46</td>
<td>4 (8.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C4/C5</td>
<td>48</td>
<td>9 (18.8)</td>
<td>46</td>
<td>6 (13.0)</td>
<td>49</td>
<td>6 (12.2)</td>
<td>48</td>
<td>10 (20.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C5/C6</td>
<td>48</td>
<td>13 (27.1)</td>
<td>46</td>
<td>13 (28.3)</td>
<td>49</td>
<td>14 (28.6)</td>
<td>48</td>
<td>16 (33.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C6/C7</td>
<td>47</td>
<td>9 (19.1)</td>
<td>46</td>
<td>8 (17.4)</td>
<td>49</td>
<td>12 (24.5)</td>
<td>48</td>
<td>12 (25.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left neural foraminal stenosis C7/T1</td>
<td>45</td>
<td>0 (0.0)</td>
<td>46</td>
<td>1 (2.2)</td>
<td>48</td>
<td>0 (0.0)</td>
<td>45</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dear Line,

You have asked the Regional Ethical Committee of Central Denmark Region whether the described project needs committee approval. After specific assessment of your study, it is not considered a health research project but rather a methodological study (following the definitions in the law of the committee §2, no. 1). Thus, the study does not require approval from the committee. Please be aware that you might need approval from the Data Protection Agency.

The law in mention is law no. 593 from June 14th 2011 about scientific ethical assessment of health research.

Kind regards,

Helle Nikkel
Secretary
Scientific Ethical Committees, Central Denmark Region
**Table A.** Overview of RCTs aimed at return to work. Workplaces involved

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Work outcomes poorer < 90 days but better > 90 days
Green: Work outcomes statistically significantly improved
Red: No difference in work outcomes between new intervention and reference intervention
LBP: Low Back Pain; MSK: Musculoskeletal pain; MHD: Mental Health Disorders
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Follow up 1 year  
*** Follow up 5 years  
**** Follow up 3 months  
***** Follow up 11 months  

Green: Work outcomes statistically significantly improved  
Red: No difference in work outcomes between new intervention and reference intervention  
LBP: Low Back Pain; MSK: Musculoskeletal pain; MHD: Mental Health Disorders