Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

van Trijffel, E.

Publication date
2015

Document Version
Final published version

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

Role and reliability of passive joint motion assessment:
Towards multivariable diagnostics and decision-making in manual therapy

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

Emiel van Trijffel

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

Emiel van Trijffel

Colophon

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

PhD thesis, Academic Medical Centre/University of Amsterdam, the Netherlands

For reasons of consistency, some terms have been standardised throughout the text. As a consequence, the text may differ in this respect from the published articles.

Printed by: Drukkerij de Gans B.V., Amersfoort, the Netherlands

ISBN: 978-90-9029304-2

Financial support for the printing of this thesis was most generously provided by Stichting Opleidingen Musculoskeletale Therapie, Amersfoort, the Netherlands.

somt
Stichting Opleidingen Musculoskeletale Therapie

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. D.C. van den Boom

ten overstaan van een door het College voor Promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 12 november 2015, te 12.00 uur

door

Emiel van Trijffel

geboren te Rotterdam

Promotiecommissie

Promotor:	Prof. Dr. P.M.M. Bossuyt	Universiteit van Amsterdam
Co-promotores:	Prof. Dr. R.A.B. Oostendorp	Radboud Universiteit
	Prof. Dr. C. Lucas	Universiteit van Amsterdam
Overige leden:	Prof. Dr. C.N. van Dijk	Universiteit van Amsterdam
	Prof. Dr. R.H. Engelbert	Universiteit van Amsterdam
	Prof. Dr. B.W. Koes	Erasmus Universiteit Rotterdam
	Prof. Dr. H.G. Schmidt	Erasmus Universiteit Rotterdam
	Prof. Dr. P. Vaes	Vrije Universiteit Brussel
	Dr. M.W. Langendam	Universiteit van Amsterdam

Faculteit der Geneeskunde

Table of contents

Chapter 1

General introduction and outline

General introduction

Manual therapy is described as a specialisation within physiotherapy that is concerned with the management of neuromusculoskeletal health problems arising from the spine and the extremities that are associated with activity limitations and participation restrictions.[1;2] Originating from early 'bone setting', chiropractic, and osteopathy, manual therapy evolved in many countries as an independent profession working alongside and in cooperation with medical physicians during the second half of the twentieth century.[3]

Manual therapy is characterised by the skill of therapists to manually evaluate joint function and to apply highly specific passive mobilising and manipulative interventions to joints.[2;4] Spinal high-velocity, low-amplitude thrust manipulation is a widely used treatment option in the US, Canada, Europe, and Australia, most commonly for neck and low-back pain.[5] It has been suggested that the mechanical stimulus induced by manual joint interventions initiates a cascade of neurophysiological responses from the peripheral and central nervous system which then produce clinical outcomes.[6] The mechanisms behind manual therapy are, however, far from established.[6] Nevertheless, there is limited evidence that spinal joint mobilisation and manipulation are effective, as well as cost-effective, in patients with non-specific neck and low-back pain although no more effective than other treatment modalities.[7-12]

This thesis is about the role and the value of passive joint motion assessment within clinical diagnostics in manual therapy. Joint function can be assessed through evaluation of a joint's mobility or stability, or through provocation of a patient's pain or other clinical symptoms. Passive manual examination of joint mobility as a diagnostic procedure has always been at the very heart of the various traditional concepts in manual therapy.[13] Such passive assessment of the quantity (e.g., range of motion, joint play, restriction) or quality (e.g., end-feel, resistance, stiffness) of joint motion is supposed to guide treatment decisions for patients.[14;15] The thesis focuses on the role and the reliability of therapists' judgements involved during passive assessment of spinal joint motion in diagnostics and decision-making in manual therapy.

The inspiration for this thesis came from my observations of my teachers, clinical mentors, and, later on, my colleagues, who showed an almost endless confidence when performing and interpreting passive intervertebral motion (PIVM) assessment of spinal motion segments. Consider the enormous difficulty involved when performing and interpreting an assessment of the amount of passive rotational motion, and the resistance perceived at the end of this range of motion in an intervertebral joint of, for instance, the lumbar spine about the size of a finger joint, which cannot be directly palpated and has a mobility of only two degrees.[16] Personally, I have never been able to demonstrate a similar level of certainty in the conclusions drawn from this testing procedure as my colleagues seemed to reach. Was I incompetent or were they overconfident?

Diagnostic test procedures should be reliable in order to yield uniform, valid, and helpful decisions about the care of patients.[17] Reliability is a component of reproducibility, along with agreement, and affects how well a test can differentiate among individuals, despite measurement errors.[18;19] At its essence, reproducibility reflects the extent to which repeated measurements in *stable* study subjects provide similar results.[19] In diagnostics, an estimate of inter-examiner reliability may then be the most clinically useful parameter to quantify to which extent two or more examiners obtain similar results when testing the same patient.[20] As such, inter-examiner reliability provides an index of a profession's diagnostic performance.[21]

Results of inter-examiner reliability studies of PIVM assessment have been mostly disappointing leading to debate within professional groups about the usefulness of testing segmental spinal mobility for decision-making in clinical practice.[22-27] However, these studies had not been systematically searched, selected, appraised, and analysed leaving uncertainty about the estimated reliability of PIVM assessment particularly as related to the studies' methodological quality. Moreover, similar conclusions could be drawn with respect to the assessment of extremity joints.

The methodology of systematic reviews of reliability studies has not yet been thoroughly developed. More specifically, no empirical evidence exists for identifying sources of bias and variation in studies of diagnostic reliability, in contrast with studies of diagnostic accuracy.[28] Such evidence would allow a more rigorous assessment of the risk of bias and concerns about the applicability of study results. As a consequence, methodological quality assessment in systematic reviews of inter-examiner reliability studies poses a challenge to reviewers. At best, reviewers inform their judgements about quality by theoretical evidence, a preliminary quality appraisal tool for studies of diagnostic reliability (QAREL), and evidence available from the context of diagnostic accuracy in which test results are verified by a reference standard.[29-31] This thesis provides early examples of the methodology proposed for systematically reviewing studies of inter-examiner reliability within the field of physical examination in musculoskeletal disorders.

As part of the methodological quality assessment of studies in systematic reviews, judgement is required in evaluating whether study results apply to the clinical setting defined in the review question. This concerns the applicability, generalisability, or external validity of a study.[32] When reviewing reliability studies of PIVM assessment, there should be a clear picture of how this procedure is used within diagnostics in clinical practice in manual therapy. It should be known to reviewers which role and position PIVM assessment takes as related to other diagnostic procedures to, for example, determine which clinical information should or should not be available to examiners before performing PIVM assessment. Most descriptions of how and when to use PIVM assessment during diagnostics stem from authority-based textbooks. It is, however, unknown how, why, and when manual therapists actually use PIVM assessment within their diagnostic strategies and clinical reasoning in patient care. This thesis describes the first practice-based exploration to understand the role and position of PIVM assessment, and how and why it is deemed important for making clinical decisions about treatment indications for patients.

Many studies of inter-examiner reliability of PIVM assessment can be classified as univariable, single-test research.[33] Such single-test research typically quantifies one test characteristic independent of any other of the test's outcomes or other clinical information. In our systematic reviews, for example, we considered studies that evaluated the inter-examiner reliability of passive joint motion assessment with respect to judgements or measurements of joint mobility only rather than those that also included other clinical findings such as provocation of pain or other clinical symptoms as outcomes of this assessment. In clinical practice in manual therapy, it is unlikely, however, that PIVM assessment is used and interpreted without other clinical findings being considered or incorporated. We propose a multivariable approach towards evaluating the process of clinical decision-making about whether or not manual spinal joint interventions are indicated, thereby recognising the dependency between all data gathered during a clinical encounter including findings from PIVM assessment.[33] Such an approach supposes knowledge of the role and position of a test within a diagnostic strategy.[34] After exploring this role and position for PIVM assessment, the thesis presents the protocol of a study to estimate the reliability among manual therapists of indicating spinal joint mobilisation or manipulation in patients with neck or low-back pain incorporating all clinical data from patient's history, observation, physical tests (including PIVM assessment), performance tests, and questionnaires.

Outline of the thesis

The objectives of this thesis were (1) to evaluate the inter-examiner reliability of passive joint motion assessment of the spine and the extremities and (2) to examine the role and position of PIVM assessment within the process of clinical reasoning and decision-making in clinical practice in manual therapy in patients with spine-related disorders.

The first two chapters describe systematic reviews of the inter-examiner reliability of passive assessment of joint motion in the spine and the extremities from a univariable, single-test research perspective. **Chapter 2** concerns the reliability of segmental PIVM assessment of the cervical (motion segments C0-T4) and lumbar (T10-S1) spine and also serves as an early attempt at exploring the methodology of systematic reviews of studies of diagnostic reliability more specifically with respect to their search strategy and methodological quality assessment. **Chapter 3** regards the inter-examiner reliability of passive assessment of joint motion in the extremities. **Chapter 3a** concerns the upper extremity (shoulder, elbow, and wrist-hand-fingers) while **Chapter 3b** relates to the lower extremity (hip, knee, and ankle-foot-toes).

In **Chapter 4**, the role and position of PIVM assessment within clinical reasoning and decision-making in clinical practice in manual therapy is explored. **Chapter 4a** presents a quantitative survey among 367 Dutch manual therapists tapping the use and interpretation of PIVM assessment as well as therapists' perceived importance and confidence regarding conclusions from this diagnostic procedure. **Chapter 4b** proceeds with a deeper exploration and understanding of the role and position of PIVM assessment within manual diagnostics using data from qualitative interviews with eight individual experts in the field and three focus groups consisting of manual therapists participating in regional consultation platforms. A model for the place of PIVM assessment within manual diagnostics is illustrated, one that integrates the theoretical concepts of professionalism and clinical reasoning.[35;36]

Within the process of manual diagnostics, **Chapter 5** focuses on the quality of biopsychosocial history taking by manual therapists in patients with (chronic) neck or low-back pain. The SCEBS method (Dutch: SCEGS methode), covering the somatic, psychological, and social dimensions of chronic pain, proves a useful starting point for the development of process indicators. The analysis of 108 patient audio recordings subsequently gives an indication of the extent of implementation of biopsychosocial history taking in Dutch manual therapy practice.

Building on the evaluations of the role and position of PIVM assessment within diagnostic strategies from previous chapters, **Chapter 6** provides a description of the protocol of a study to estimate the inter-examiner reliability among manual therapists of indicating spinal joint mobilisation and manipulation in patients with neck or low-back pain. This chapter proposes a multivariable approach to investigating reliability. Instead of evaluating the inter-examiner reliability of independent single tests or test outcomes, it focuses on the reliability of the decision about whether or not manual spinal joint interventions are indicated, thereby integrating clinical data from a full diagnostic process including judgements from PIVM assessment. As such, it is suggested as an initial step towards an alternative approach to the currently popular prediction rules and other classification systems for identifying those patients with spinal disorders that may show a better response to manual therapy.

Chapter 7 presents a series of basic scientific experiments investigating biomechanical effects of passive movements in joints as a function of time. It reports on the first attempts to visualise and measure *in vivo* time-dependent changes in synovial fluid volume after passive joint motion assessment, mobilisation, and high-velocity, low-amplitude thrust manipulation using ultrasonography and magnetic resonance imaging. Data from these experiments could be of importance to researchers in the field of reliability of motion assessment who need to minimise their study's risk of bias by ensuring stability of the participants' characteristic under study, i.e., their joint's mobility, during the research.

The thesis concludes with a summary in **Chapter 8** which also contains a general discussion of the thesis's findings and subsequent directives for clinical practice, research, and education in manual therapy.

References

1. NVMT (Dutch Association for Manual Therapy). Professional competence profile Manual therapist (October 2014). Amersfoort, The Netherlands. Available at: http://nvmt.fysionet.nl/diensten/richtlijnen_publicaties-en-bcp.html.pdf. [In Dutch] (accessed 12 October, 2014)
2. IFOMPT (International Federation of Orthopaedic Manipulative Physical Therapists). Educational standards in orthopaedic manipulative therapy. Part A: Education. Accepted 11 June 2008. Amended September 2013. Available at: http://www.ifompt.com/Educational+Standards/Standards+Document.html (accessed 12 October, 2014)
3. Pettman E. A history of manipulative therapy. J Man Manip Ther 2007;15:165-74.
4. Van Ravensberg CDD, Oostendorp RAB, Van Berkel LM, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual therapy: Differences in patient characteristics. J Man Manip Ther 2005;13:113-24.
5. Hurwitz EL. Epidemiology: Spinal manipulation utilization. J Electromyogr Kinesiol 2012;22:648-54.
6. Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: A comprehensive model. Man Ther 2009;14:531-8.
7. Bronfort G, Haas M, Evans R, Leininger B, Triano J. Effectiveness of manual therapies: The UK evidence report. Chiropr Osteopat 2010;18:3.
8. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Brønfort G, Hoving JL. Manipulation or mobilisation for neck pain. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD004249.
9. Rubinstein SM, Van Middelkoop M, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for chronic low-back pain. Cochrane Database of Systematic Reviews 2011, Issue 2. Art. No.: CD008112.
10. Rubinstein SM, Terwee CB, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for acute low-back pain. Cochrane Database of Systematic Reviews 2012, Issue 9. Art. No.: CD008880.

11. Michaleff ZA, Lin C-WC, Maher CG, Van Tulder MW. Spinal manipulation epidemiology: Systematic review of cost-effectiveness studies. J Electromyogr Kinesiol 2012;22:655-62.
12. Tserstvadze A, Clar C, Court R, Clarke A, Mistry H, Sutcliffe P. Cost-effectiveness of manual therapy for the management of musculoskeletal conditions: A systematic review and narrative synthesis of evidence from randomized controlled trials. J Manipulative Physiol Ther 2014;37:343-62.
13. Farrell JP, Jensen GM. Manual therapy: A critical assessment of role in the profession of physical therapy. Phys Ther 1992;72:843-52.
14. Jull G, Treleaven J, Versace G. Manual examination: Is pain provocation a major cue for spinal dysfunction? Aust J Physiother 1994;40:159-65.
15. Bialosky JE, Simon CB, Bishop MD, George SZ. Basis for spinal manipulative therapy: A physical therapist perspective. J Electromyogr Kinesiol 2012;22:643-7.
16. Pearcy MJ, Tibrewal SB. Axial rotation and lateral bending in the normal lumbar spine measured by three-dimensional radiography. Spine 1984;9:582-7.
17. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
18. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (4th edition). Oxford, UK: Oxford University Press, 2008.
19. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
20. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, Norman GR, Guyatt GH. Evaluating agreement: Conducting a reliability study. J Bone Joint Surg Am 2009;91:99-106.
21. Haas M. How to evaluate intraexaminer reliability using an interexaminer reliability study design. J Manipulative Physiol Ther 1995;18:10-5.
22. Breen A. The reliability of palpation and other diagnostic methods. J Manipulative Physiol Ther 1992;15:54-6.
23. Maher C, Latimer J. Pain or resistance – the manual therapists' dilemma. Aust J Physiother 1992;38:257-60.
24. Troyanovich SJ, Harrison DD. Motion palpation: It's time to accept the evidence. J Manipulative Physiol Ther 1998;21:568-71.

25. Huijbregts PA. Spinal motion palpation: A review of reliability studies. J Man Manip Ther 2002;10:24-39.
26. Haas M, Groupp E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay assessment as an indicator for spinal manipulation. Spine 2003;28:1091-6.
27. Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: An examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehabil 2005;86:1745-52.
28. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, and the QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol 2013;66:1093-1104.
29. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010;63:854-61.
30. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology 2003;3:25.
31. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, and the QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155;529-36.
32. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" Lancet 2005;365:82-93.
33. Moons KG, Biesheuvel CJ, Grobbee DJ. Test research versus diagnostic research. Clin Chem 2004;50:473-6.
34. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089-92.
35. Freidson E. Professionalism. The third logic. Oxford, UK: Blackwell Publishers Ltd, 2001.
36. Jones MA, Jensen G, Edwards I. Clinical reasoning in physiotherapy. In: Higgs J, Jones MA, Loftus S, Christensen N (editors). Clinical reasoning in the health professions (3rd edition). Edinburgh, UK: Elsevier/Butterworth Heinemann, 2008.

Chapter 2

Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review

Emiel van Trijffel, Quirine Anderegg, Patrick Bossuyt, Cees Lucas

Man Ther 2005;10:256-69

Abstract

Background: Passive intervertebral motion (PIVM) assessment of the spine is used to decide on treatments for patients with neck or low-back pain. Inter-examiner reliability has been a matter of debate resulting in questions about professional credibility and accountability.

Methods: A systematic review was conducted to determine the inter-examiner reliability of segmental PIVM assessment in the cervical (motion segments C0-T4) and lumbar (T10-S1) spine as well as to explore sources of heterogeneity. A structured search for relevant studies in MEDLINE and CINAHL published up to March 31, 2004, was followed by extensive reference tracing and hand searching. Studies presenting estimates of reliability for individual motion segments were included. No language restrictions were imposed. Study quality was assessed using criteria derived from the Standards for Reporting of Diagnostic Accuracy (STARD) statement and a quality assessment tool for studies of diagnostic accuracy included in systematic reviews (QUADAS). Study selection, quality assessment, and data extraction were performed by two reviewers independently. Qualitative analyses and additional subgroup analyses were conducted.

Results: Nineteen studies were included. Overall, the inter-examiner reliability of PIVM assessment was poor to fair. Two studies satisfied the criteria for external and internal validity of which one found fair to moderate reliability. Assessment of motion segments C1-C2 and C2-C3 almost consistently reached at least fair reliability.

Conclusions: The inter-examiner reliability of PIVM assessment of the cervical and lumbar spine was poor. However, most studies were found to be of low methodological quality. We propose explicit recommendations for the conduct and reporting of future research.

Introduction

An overview of epidemiologic research has shown high prevalence rates of neck and low-back pain in developed countries.[1] Generally, at some point during the clinical course, many patients suffering from these conditions are treated by manual practitioners such as physiotherapists, manual therapists, chiropractors, osteopaths, and physicians. In the Netherlands, according to guidelines, general practitioners may refer patients with low-back pain persisting longer than six weeks.[2] Fifty-nine percent of patients with chronic neck pain are referred to a physiotherapist or manual therapist.[3]

Passive assessment of the quantity and quality of motion - also known as motion palpation - in individual vertebral motion segments guides decisions on treatment.[4] Reliability reflects the extent to which practitioners are able to differentiate diagnostically among individuals who vary in characteristics.[5] Furthermore, an estimate of inter-examiner reliability can be used to quantify the extent to which practitioners show variability in diagnostic assessment.[6] A satisfactory level of inter-examiner reliability is a prerequisite for valid and uniform decisions about patients.[7] Variability among examiners has empirically been shown to affect diagnostic accuracy.[8] At this moment, it is unclear to what extent practitioners vary in their motion assessment of the spine.

Inter-examiner reliability of passive intervertebral motion (PIVM) assessment of the spine has been a matter of debate resulting in questions about professional credibility and accountability.[9-11] Four narrative reviews concerning reliability of spinal motion assessment have been published of which two dealt with the lumbar spine only.[12-15] None of these reviews formally assessed the methodological quality of included studies. Two extensive systematic reviews have appeared covering the reliability of chiropractic tests for the lumbo-pelvic spine and spinal palpation tests.[16;17] In both reviews, it was concluded that the inter-examiner reliability of PIVM assessment was low. Seffinger et al[17] added that assessing regional range of motion was more reliable than evaluating segmental range of motion. However, in both reviews, the criteria for assessing methodological quality of studies were not substantiated by evidence of variation and bias in diagnostic research. Furthermore, none of all the above mentioned reviews explicitly analysed reliability for individual motion segments. So far, no single study has been able to demonstrate acceptable inter-examiner reliability for PIVM assessment. A systematic review on this topic is needed to allow for an objective appraisal of existing evidence.[18]

We conducted a systematic review of the available literature to determine the inter-examiner reliability of segmental PIVM assessment of the cervical and lumbar spine. In addition, we explored sources of heterogeneity.

Methods

Study selection

Assisted by a clinical librarian, we developed a structured search strategy to identify relevant studies in the MEDLINE database (through PubMed) published between January 1, 1966, and March 31, 2004 (Box 1).

Box 1. Search strategy for MEDLINE (PubMed) using text words [tw] and Medical Subject Headings [mh]

(zygapophyseal joint[mh] OR spine[mh] OR spine[tw] OR spinal[tw] OR lumbal[tw] OR lumbar[tw] OR lumbosacr*[tw] OR cervical[tw] OR back[mh] OR back[tw] OR neck[mh] OR neck[tw])

AND

(motion[mh] OR range of motion, articular[mh] OR motion*[tw] OR ((movement*[tw] OR mobility[tw]) AND (manual*[tw] OR palpat*[tw] OR passive[tw] OR intersegment*[tw])) OR joint-play[tw] OR joint play[tw])

AND

(observer variation[mh] OR ((reliability[tw] OR reproducibility of results[mh] OR reproducibility[tw] OR concordance[tw] OR repeatability[tw] OR agreement[tw] OR variation*[tw] OR variabilit*[tw])

AND

(interexaminer[tw] OR interobserver[tw] OR interrater[tw] OR intertester[tw] OR examiner*[tw] OR observer*[tw] OR rater*[tw] OR tester*[tw])))

The search and study selection were performed by two reviewers (EvT, QA) independently. Based on information in title and abstract, possibly relevant studies were selected and retrieved as a full article. Studies, or subsets of studies, meeting the following criteria were included:

- published as a full article;
- using a repeated-measures, inter-examiner reliability design;
- evaluating passive motion assessment of one or more motion segments of the cervical (C0-T4) and lumbar (T12-S1) spine;

- applying judgement criteria that could either concern the quantity (e.g., range of motion, joint play, restriction) or quality (e.g., end-feel, resistance, stiffness) of motion;
- presenting estimates of inter-examiner reliability for individual motion segments.

No restrictions were imposed on language and date of publication. Abstracts and theses were not included. Studies evaluating active movements or incorporating other clinical symptoms such as pain into the judgement process were not considered.

The first reviewer performed an additional search in the CINAHL database (1982 - March 31, 2004). All of the retrieved article references and relevant reviews were further examined by the first reviewer for additional publications. This strategy was complemented by hand searching of nine journals (January 1, 1990, to March 31, 2004). A complete list of journals is available from the authors. Eligibility was checked by the second reviewer. Disagreements were resolved by discussion. If disagreement persisted, the judgement of a third reviewer (CL) was decisive.

Quality assessment

A validated list of criteria for assessing the methodological quality of inter-examiner reliability studies was not available. We therefore developed a list of 11 criteria for assessing study quality (Box 2).

Seven of the criteria were derived from evidence of variation and design-related bias in diagnostic accuracy studies, the Standards for Reporting of Diagnostic Accuracy (STARD) statement, and a validated tool for assessing quality of studies of diagnostic accuracy included in systematic reviews (QUADAS).[8;19-22] Based on theoretical evidence, Criteria 5, 9, 10, and 11 were added to fit the context of reliability.[5;6;23-30] These criteria were designed to tap the domains of external validity (Criteria 1-3), internal validity (Criteria 4-8), and statistical methods used (Criteria 9-11). The scores on Criteria 4 and 7 were assumed to be of decisive importance for internal validity. After a training session, two papers were used to evaluate the interpretability and applicability of criteria by all reviewers.[31;32]

Box 2. Criteria for assessing methodological quality

1. Was a representative sample of participants used?
2. Was a representative sample of examiners used?
3. Is replication of the assessment procedure possible?
4. Were participants' characteristics under study stable during research?
5. Was an estimate of intra-examiner reliability sufficiently large?
6. Were examiners blinded to clinical information from participants?
7. Were examiners blinded to each other's results?
8. Can non-random loss to follow-up be ruled out?
9. Were appropriate measures used for calculating reliability?
10. Can prevalence bias be ruled out?
11. Can systematic bias be ruled out?

EvT and QA, who were not blinded to information on authors and journals, independently assessed the methodological quality of all included studies. Criteria were scored by answering with "Yes", "No", or "?" (unclear because of insufficient information). Criteria were equally weighted. Inter-reviewer reliability was analysed by calculating percentage agreement and Cohen's kappa (κ). Disagreements were resolved by discussion. In case disagreement persisted, CL made the final decision.

Data extraction

We extracted data from the original studies on participants (number, age, gender, clinical characteristics, setting), examiners (number, profession, expertise, pre-training, experience), assessment procedure (subject position, motion segments, motion directions), judgement criteria and scales (quantitative and qualitative classifications), and inter-examiner reliability for individual motion segments (point estimates and estimates of precision). EvT and QA extracted data independently. If disagreement persisted after discussion, consensus was met consulting CL.

Data analysis

Qualitative analyses were conducted by examining results on reliability from studies with high methodological quality, as well as by examining characteristics of studies that showed the highest and lowest levels of reliability. Additionally, analyses for subgroups of participants, examiners, assessment procedures, judgement criteria and scales, and motion segments were performed. Analyses were carried out for the cervical spine and lumbar spine separately. Value labels for corresponding ranges of Cohen's kappa statistics were used as assigned by Landis & Koch (Box 3).[33]

Box 3. Value labels for ranges of Cohen's kappa (κ)

κ < 0.00	Poor
κ 0.00-0.20	Slight
κ 0.21-0.40	Fair
κ 0.41-0.60	Moderate
κ 0.61-0.80	Substantial
κ 0.81-1.00	Almost perfect

Intraclass correlation coefficients (ICC) >0.75 are assumed to indicate an acceptable level of reliability.[34]

Results

Flow of studies through the review

Searching MEDLINE yielded 228 hits. Of these, 18 possibly relevant studies were retrieved as a full article.[35-52] Eight studies fulfilled all eligibility criteria. Searching CINAHL (208 hits) led to the inclusion of one more study.[53] Reference tracing and hand searching yielded another 16 possibly relevant studies[54-69] of which eight met the inclusion criteria. Thirteen studies were excluded for reasons of design features[54], evaluating active movements[38;43;49], incorporating other clinical examination symptoms into the judgement process[42;45;46;66;68;69], and not examining individual motion segments.[35;37;40] Five studies did not present estimates of inter-examiner reliability for individual motion segments.[50;58;60;64;65] First authors were contacted and requested to provide segmental data. As a result, another two studies could be included.[50;65] In total, 16 studies were excluded while 19 studies could be included in this review (Figure). There were no disagreements between reviewers on the selection of studies.

Characteristics of included studies

Nine studies[36;44;47;48;50;52;56;65;67] examined reliability for the cervical spine while 10[39;41;51;53;55;57;59;61-63] evaluated the lumbar spine. Study characteristics are given in Table 1 (cervical spine) and Table 2 (lumbar spine).

In seven studies[44;48;51;53;62;63;67], physiotherapists specialising in manual therapy acted as examiners and two studies[47;61] used physicians specialised in manual medicine. There were no disagreements between reviewers on the extraction of data.

Quality of studies

Methodological quality scores of included studies are presented in Table 3.

Three studies[48;51;67] satisfied both decisive criteria for internal validity of which two[48;51] also met all criteria for external validity. There were 10 disagreements between reviewers on quality scores resulting in 95% agreement and an inter-reviewer reliability (κ) of 0.93. All disagreements were resolved by discussion, consequently there was no need consulting the third reviewer for a final decision.

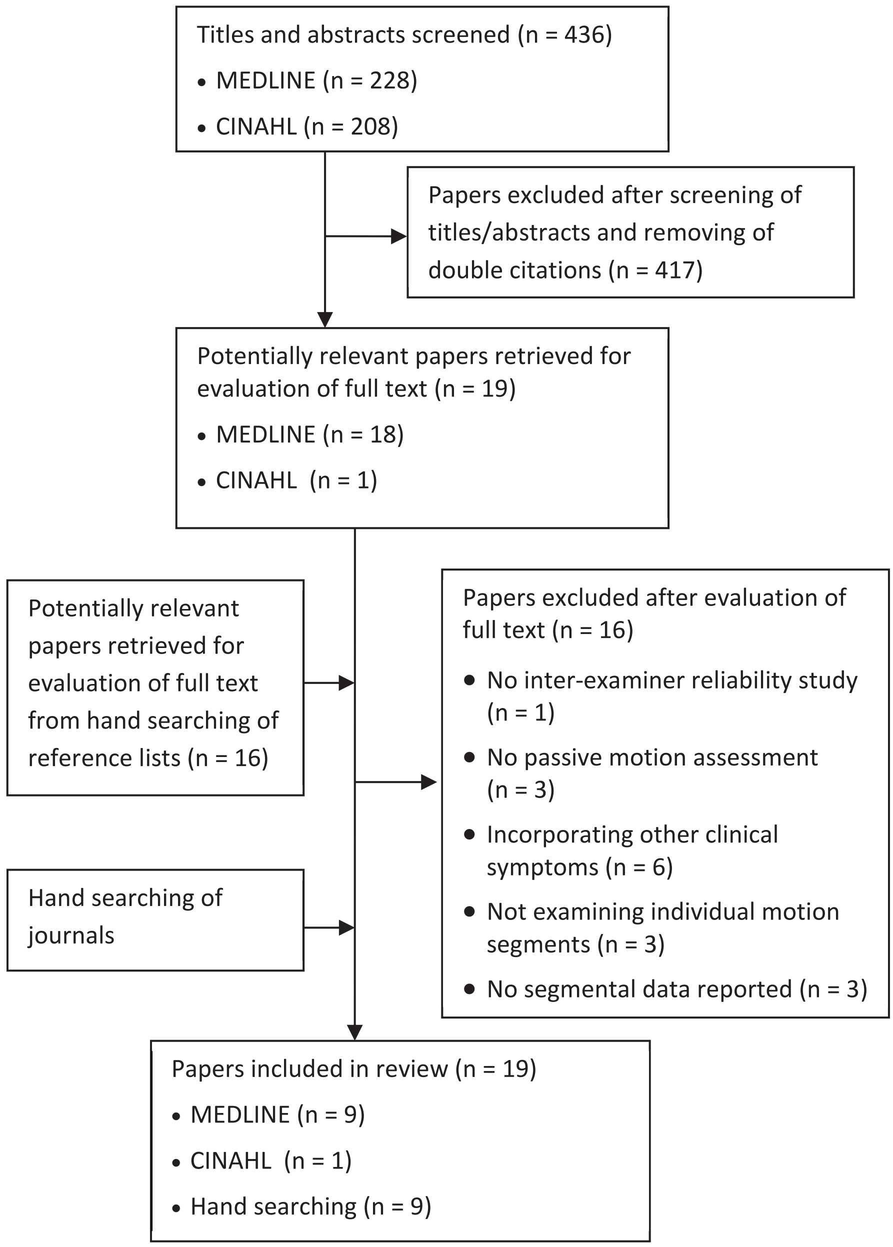

Figure. Flow of studies through the review

Inter-examiner reliability by region

Cervical spine (n = 9)

Data on estimates of inter-examiner reliability of PIVM assessment are given in Table 1 (last column). Inter-examiner reliability for the cervical spine ranged from poor to substantial. Overall, reliability was poor to fair.

Table 1. Characteristics of included studies: cervical spine *(n = 9)*

Study	Participants	Examiners	Assessment procedure	Judgement criteria and scales	Inter-examiner reliability
Christensen[50]	29 patients referred to cardiology with known or suspected stable angina pectoris + 27 controls referred to nuclear medicine. Taken from original sample of 107 with age range 31-74 yrs and 68 (64%) males. University Hospital, Denmark.	2 pre-trained chiropractors.	*Subject seated*: MSs T1-T4 in lateral flexion R/L, rotation R/L. *Subject prone*: MSs T1-T4 joint play. (reference cited)	*Abnormality* (based on end-play restriction and joint play, respectively): absent-present.	*Seated*: <u>T1-T2</u> PA 75%, κ 0.11 <u>T2-T3</u> PA 77%, κ 0.00 <u>T3-T4</u> PA 75%, κ -0.32 *Prone*: <u>T1-T2</u> PA 75%, κ -0.19 <u>T2-T3</u> PA 73%, κ -0.42 <u>T3-T4</u> PA 68%, κ -0.23
DeBoer[36]	40 healthy students of chiropractic. Mean age 26.2 yrs (range 21-44). 40 (100%) males. College of Chiropractic, US.	3 (3 pairs) chiropractors. Range 5-14 yrs of experience.	*Subject seated*: Vertebrae C1-C7 in flexion, extension, rotation, lateral flexion R/L. (reference cited)	*Fixation*: normal-slight-obvious.	<u>C1-C2</u> (pooled data) PA 56%, κ_w 0.23; PA 21%, κ_w -0.03; PA 38%, κ_w 0.09 <u>C6-C7</u> (pooled data) PA 44%, κ_w 0.40; PA 58%, κ_w 0.41; PA 49%, κ_w 0.45
Fjellner[44]	47 healthy volunteers by advertising or inquiry. Mean age 37.9 yrs (SD ±9.5, range 18-63). 8 (17%) males. Sweden.	2 physiotherapists specialised in orthopaedic manual therapy. 6 and 12 yrs of experience.	*Subject seated*: MS C0-C1 in flexion, extension. MS C1-C2 in rotation R/L. MSs C2-T4 in rotation R/L. First rib R/L. *Subject supine*: MSs C2-T4 in flexion, extension. *Subject left side-lying*: MSs C2-T4 joint play. (references cited)	*Range of motion* MSs C0-C1/C1-C2/ C2-T4: reduced-normal-increased, first rib: reduced-normal. *Joint play* MSs C2-T4: reduced-normal-increased. *End-feel* MSs C0-C1/ C1-C2: hard-normal-empty.	*Range of motion*: <u>C0-C1</u> flexion PA 62%, κ_w 0.00 (CI [-0.27, 0.27]), extension PA 87%, κ_w NC; <u>C1-C2</u> rotation R PA 62%, κ_w 0.15 (CI [-0.14,0.44]), L PA 79%, κ_w 0.41 (CI [0.096, 0.72]); <u>C2-T4</u> ranging from flexion <u>C7-T1</u> PA 72%, κ_w -0.16 (CI [-0.26,-0.062]) to rotation L <u>T2-T3</u> PA 81%, κ_w 0.49 (CI [-0.22,0.76]); <u>first rib</u> R PA 92%, κ_w NC, L PA 77%, κ_w 0.06 (CI [-0.21,0.33]) *Joint play*: ranging from <u>C4-C5</u> PA 79%, κ_w -0.05 (CI [-0.089,-0.011]) to <u>C3-C4</u> PA 83%, κ_w 0.36 (CI [0.27,0.69]) *End-feel*: <u>C0-C1</u> flexion PA 64%, κ_w 0.01 (CI [-0.24,0.26]), extension PA 87%, κ_w NC; <u>C1-C2</u> rotation R PA 60%, κ_w 0.06 (CI [-0.21,0.33]), L PA 75%, κ_w 0.18 (CI [-0.075,0.43])

Haas[65]	73 first year students of chiropractic of which 48 (66%) mild symptomatic. Mean age 27.1 yrs (SD ±5.2). 49 (67%) males. College of Chiropractic, US.	2 pre-trained chiropractors (faculty members). 15 yrs of experience.	*Subject seated*: MS T3-T4 in rotation R/L. (reference cited)	*Restriction* (based on hard end-play): absent-present.	T3-T4 rotation R κ -0.03 (SE 0.01), L κ NC, either R/L κ -0.04 (SE 0.03)
Mior[56]	59 healthy students of chiropractic. Age range 22-30 yrs. College of Chiropractic, Canada.	2 pre-trained students of chiropractic (first year of clinical training).	*Subject supine*: Vertebra C1 in lateral flexion R/L, anterior rotation.	*Fixation* (based on joint play and end-feel): absent-present.	C1 PA 62%, κ 0.15
Pool[52]	32 patients with neck complaints. Mean NDI score 15.2 (SD ±8.3), 56.3% had previous episodes, mean present pain on 11-point scale 4.2 (SD ±2.3), median duration of pain 13.5 wks. Mean age 45.5 yrs (SD ±9.2). 12 (37.5%) males. Practice for Physical and Manual Therapy, the Netherlands.	2 pre-trained physical therapists.	*Subject supine*: MS C0-C1in flexion. MS C1-C2 in rotation R/L. MSs C2-T2 in lateral flexion R/L.	*Limitation of movement* (based on range of motion and resistance): yes-no.	C0-C1 flexion PA 77%, κ 0.29 C1-C2 rotation R PA 84%, κ 0.20, L PA 90%, κ 0.37 C2-T2 ranging from C4-C5 lateral flexion L PA 68%, κ -0.09 to C2-C3 lateral flexion L PA 84%, κ 0.63
Schöps[47]	20 patients with cervical spine syndrome: mean age 37 yrs (range 21-55), 8 (40%) males + 20 healthy volunteers: mean age 33 yrs (range 20-49), 10 (50%) males. Clinic for Physical Medicine and Rehabilitation, Germany.	5 physicians specialised in manual medicine.	MS C0-C1 in lateral flexion R/L, rotation R/L. MS C1-C2 in nodding-flexion-rotation R/L. MS C2-C3 in nodding-flexion-lateral flexion R/L. MSs C3-C6 in unspecified direction R/L. MSs C6-T1 in rotation R/L.	*Hypomobility*: absent-present.	C0-C1 lateral flexion R κ 0.04, L κ 0.13; rotation R κ 0.08, L κ 0.04 C1-C2 R κ 0.22, L κ 0.28 C2-C3 R κ 0.04, L κ 0.34 C3-C4 R κ 0.43, L κ 0.06 C4-C5 R κ 0.03, L κ 0.03 C5-C6 R κ 0.28, L κ 0.17 C6-C7 R κ 0.44, L κ 0.29 C7-T1 R κ 0.26, L κ 0.30

Smedmark[48]	61 patients seeking care for non-specific neck problems. Age range 20-71 yrs. 15 (24.5%) males. Private Clinic, Sweden.	2 pre-trained physical therapists specialised in orthopaedic manipulative therapy. Over 25 yrs of experience.	*Subject seated*: MS C1-C2 in rotation R/L. *Subject supine*: MS C2-C3 in lateral flexion R/L. First rib R/L. *Subject side-lying*: MS C7-T1 in flexion, extension.	*Stiffness* (based on range of motion and end-feel): yes-no difference when R compared to L. For C7-T1 compared to C6-C7 and T1-T2.	C1-C2 PA 87%, κ 0.28 C2-C3 PA 70%, κ 0.43 C7-T1 PA 79%, κ 0.36 first rib PA 70%, κ 0.35
Strender[67]	50 volunteers of which 25 with complaints in neck-shoulder region. Mean age 41.7 yrs (SD ±10.4, range 21-66). 13 (26%) males. Sweden.	2 pre-trained physiotherapists specialised in manual medicine. 21 and 23 yrs of experience.	*Subject supine*: MS C0-C1 in lateral flexion–rotation R/L. MSs C0-C2 in rotation (in flexion-position) R/L. MS C2-C3 in lateral flexion R/L. (reference cited)	*Mobility*: yes-no difference when R and L compared.	C0-C1 PA 26%, κ 0.091 (CI [-0.22,0.40]) C0-C2 PA 42.9%, κ 0.15 (CI [-0.06,0.37]) C2-C3 PA 44%, κ 0.057 (CI [-0.23,0.35])

CI: 95% confidence interval, κ: kappa, $κ_w$: weighted kappa, L: left, MS: motion segment, NC: not calculated, NDI: Neck Disability Index, PA: percentage agreement, R: right

Table 2. Characteristics of studies: lumbar spine *(n = 10)*

Study	Participants	Examiners	Assessment procedure	Judgement criteria and scales	Inter-examiner reliability
Bergström[57]	100 healthy students of chiropractic. UK.	2 students of chiropractic.	*Subject seated*: Vertebrae L1-L5 in lateral flexion R/L.	*Fixation* (based on hard end-feel): absent-present.	L1 R PA 83%, L PA 75% L2 R PA 82%, L PA 79% L3 R PA 88%, L PA 84% L4 R PA 87%, L PA 89% L5 R PA 78%, L PA 73%
Boline[59]	23 symptomatic LBP patients + 27 asymptomatic. 27 (54%) males. US.	2 pre-trained chiropractors (members of campus Motion Palpation Club). One senior intern and one recent graduate.	*Subject seated*: MSs T12-S1 in flexion, extension, lateral flexion R/L, rotation R/L.	*Fixation* (based on hard end-feel): normal-obvious.	N = 50 / N = 23 T12-L1 PA 70%, κ 0.31 / PA 65%, κ 0.32 L1-L2 PA 60%, κ -0.02 / PA 61%, κ 0.05 L2-L3 PA 74%, κ -0.02 / PA 74%, κ 0.09 L3-L4 PA 78%, κ 0.31 / PA 87%, κ 0.33 L4-L5 PA 82%, κ 0.19 / PA 87%, κ 0.25 L5-S1 PA 90%, κ -0.05 / PA 87%, κ -0.06
Gonella[55]	5 healthy students of physical therapy. Age range 22-27 yrs. 0 (0%) males. US.	5 pre-trained physical therapists. 3, 3, 4, 5 and 20 yrs of experience.	MSs T12- S1 in flexion, lateral flexion R/L, rotation R/L.	*Mobility*: 7-point scale (with half points) ranging from 0=ankylosed to 6=unstable with reference point the expected normal for age, body type and activity level.	Only descriptive statistics averaged over all motion directions: mean (SD) T12-L1 ranging from 2.93 (0.18) to 3.23 (0.41) L1-L2 ranging from 2.80 (0.30) to 3.00 (0.00 and 0.28) L2-L3 ranging from 2.60 (0.35) to 2.80 (0.30) L3-L4 ranging from 2.05 (0.51) to 2.85 (0.76) L4-L5 ranging from 2.18 (0.41) to 2.73 (0.26) L5-S1 ranging from 2.23 (0.34) to 3.00 (0.36)
Hicks[51]	63 patients with current complaints of LBP recruited either as consecutive participants in research on LBP or as patients referred to an outpatient physical therapy clinic. 51 (80.9%) had previous episodes, mean Oswestry score 17.8 (SD ±11.3, range 92-52). Mean	4 pre-trained examiners of which 3 physical therapists and 1 physical therapist/chiropractor. Specialised in orthopaedic physical therapy (2) and experienced in an orthopaedic setting (2). 4, 5, 6 and 8 years of experience.	*Subject prone*: Vertebrae L1-L5 in antero-posterior direction by applying an anteriorly directed pressure on spinous process. (reference cited)	*Mobility*: hypermobile-normal-hypomobile relative to adjacent motion segments and expectation of the examiner.	L1 PA 68%, κ_w 0.26 (CI [-0.01,0.53]) L2 PA 69%, κ_w 0.17 (CI [-0.13, 0.47]) L3 PA 52%, κ_w -0.02 (CI [-0.25,0.28]) L4 PA 58%, κ_w 0.11, (CI [-0.26,0.35]) L5 PA 65%, κ_w 0.18 (CI [-0.03, 0.49])

	age 36.0 yrs (SD ±10.3, range 20-66). 25 (39.6%) males. US.				
Inscoe[63]	6 volunteers currently experiencing LBP (but have not sought care) and a reported history of 2 or more previous episodes. Mean age 29.3 yrs (range 24-34). 2 (33.3%) males. US.	2 physical therapists specialised in orthopaedic manual therapy. 4-5 yrs of experience.	*Subject right side-lying*: MSs T12-S1 in flexion with double leg flexion technique. (reference cited)	*Mobility*: normal-hypomobile-hypermobile relative to the expected normal for age, body type, gender and activity level.	T12-L1 PA 33.33% L1-L2 PA 58.33% L2-L3 PA 50.0% L3-L4 PA 41.67% L4-L5 PA 58.33% L5-S1 PA 50.0%
Keating[41]	21 LBP patients + 25 asymptomatic students. Age range 23-60 yrs. 20 (43.5%) males. US.	3 (3 pairs) pre-trained chiropractors. 2.5, 5 and 10 yrs of experience.	*Subject seated*: MSs T11-S1 in flexion, extension, lateral flexion R/L, rotation R/L. (reference cited)	*Fixation* (based on hard end-feel): absent-present.	T11-T12 κ -0.04; κ 0.03; κ -0.04 (mean κ -0.02) T12-L1 κ 0.09; κ -0.15; κ 0.00 (mean κ -0.02) L1-L2 κ 0.23; κ -0.13; κ 0.01 (mean κ 0.04) L2-L3 κ 0.14; κ -0.14; κ 0.25 (mean κ 0.08) L3-L4 κ 0.13; κ -0.18; κ -0.04 (mean κ -0.03) L4-L5 κ 0.09; κ 0.29; κ 0.28 (mean κ 0.22) L5-S1 κ 0.31; κ 0.22; κ 0.17 (mean κ 0.23)
Maher[62]	90 patients with non-specific mechanical LBP. 82% previous history, mean time since onset 45.2 days (SD ±100.0, range 1-730). Mean age 45.37 yrs (SD ±14.16, range 21-78). 34 (37.7%) males. Physical Therapy Clinics, Australia.	6 (3 pairs) physical therapists specialised in manipulative physiotherapy. Range 8-21 yrs of experience.	*Subject prone*: Vertebrae L1-L5 in antero-posterior direction by applying an anteriorly directed force over spinous process. (reference cited)	*Stiffness*: 11-point scale ranging from -5=markedly decreased stiffness to 5=markedly increased stiffness with 0=normal stiffness relative to the expected normal.	L1 PA 20%, ICC 0.32; PA 33%, ICC 0.38; PA 33%, ICC -0.14 L2 PA 20%, ICC 0.30; PA 20%, ICC 0.15; PA 23%, ICC -0.40 L3 PA 40%, ICC 0.18; PA 13%, ICC 0.28; PA 27%, ICC -0.25 L4 PA 27%, ICC 0.41; PA 26%, ICC 0.54; PA 30%, ICC 0.00 L5 PA 43%, ICC 0.73; PA 20%, ICC 0.37; PA 23%, ICC -0.25
Mootz[39]	60 students of chiropractic. US.	2 pre-trained chiropractors. 7 and 10 yrs of experience.	*Subject seated*: MSs L1-S1 in flexion, extension, lateral flexion R/L, rotation R/L. 2 sessions. (reference cited)	*Fixation* (based on hard end-feel): absent-present.	L1-L2 PA 80%, κ -0.06; PA 85%, κ -0.05 L2-L3 PA 76.7%, κ -0.13; PA 85%, κ 0.11 L3-L4 PA 70%, κ -0.17; PA 75%, κ -0.03 L4-L5 PA 63.3%, κ -0.02; PA 61.7%, κ 0.08 L5-S1 PA 73.3%, κ 0.17; PA 73.3%, κ 0.08

Richter[61]	35 patients with deep back pain. Rehabilitation Clinic, Germany.	5 physicians specialised in manual medicine.	*Subject seated*: Vertebrae L1-L5 in flexion, extension, lateral flexion R/L, rotation R/L. *Subject prone*: Vertebrae L1-L5 in antero-posterior direction. (reference cited)	*Mobility*: hypermobile-normal-hypomobile.	*Seated*: L1 ranging from flexion κ 0.18 to lateral flexion L κ 0.72 L2 ranging from flexion κ 0.20 to lateral flexion L κ 0.69 L3 ranging from lateral flexion L κ 0.25 to lateral flexion R κ 0.34 L4 ranging from lateral flexion R κ 0.11 to rotation L κ 0.29 L5 ranging from lateral flexion R κ 0.08 to flexion κ 0.25 *Prone*: L1 κ 0.14, L2 κ 0.18, L3 κ 0.11, L4 κ 0.08, L5 κ 0.17
Strender[53]	Physiotherapists' group: 50 patients. Mean age 37.7 yrs (SD ±11.7, range 16-69). 17 (34%) males. Physicians' group: 21 patients. Mean age 41.2 yrs (SD ±15.7, range 20-71). 11 (52.4%) males. Private Outpatient Clinic specializing in back pain, Sweden.	4 pre-trained of which 2 physiotherapists specialised in manual medicine and 2 physicians.	*Subject side-lying*: MSs L4-S1 in angulair and translational directions with hips and knees flexed. (reference cited)	*Mobility*: decreased-normal-increased.	**physiotherapists** L4-L5 PA 82%, κ_w 0.66 (CI [0.45,0.86]) L5-S1 PA 80%, κ_w 0.75 (CI [0.60,0.90])

CI: 95% confidence interval, κ: kappa, κ_W: weighted kappa, LBP: low-back pain, L: left, MS: motion segment, PA: percentage agreement, R: right

Table 3. Quality of studies *(n = 19)*

	External validity			Internal validity					Statistical methods		
Study	1	2	3	4	5	6	7	8	9	10	11
Cervical spine											
Christensen[50]	N	?	Y	?	N	Y	Y	N	Y	N	Y
DeBoer[36]	N	Y	Y	?	N	?	Y	Y	Y	?	?
Fjellner[44]	N	Y	Y	N	?	N	Y	Y	?	?	Y
Haas [65]	N	N	Y	?	?	?	Y	Y	Y	?	?
Mior[56]	N	N	Y	Y	N	?	?	Y	Y	Y	Y
Pool[52]	Y	?	N	?	?	?	Y	Y	Y	?	?
Schöps[47]	N	?	N	?	?	Y	?	Y	Y	?	?
Smedmark[48]	Y	Y	Y	Y	?	?	Y	Y	Y	?	?
Strender[67]	N	Y	Y	Y	?	N	Y	Y	Y	N	N
Lumbar spine											
Bergström[57]	N	N	Y	?	N	?	N	Y	?	?	?
Boline[59]	?	Y	N	?	?	N	Y	Y	Y	?	?
Gonella[55]	N	Y	N	?	N	?	?	Y	N	?	?
Hicks[51]	Y	Y	Y	Y	?	?	Y	Y	?	?	N
Inscoe[63]	N	Y	Y	N	N	N	Y	Y	?	?	?
Keating[41]	N	Y	Y	?	?	?	Y	Y	Y	?	?
Maher[62]	Y	Y	Y	?	?	N	Y	Y	?	?	?
Mootz[39]	N	Y	Y	?	N	?	Y	Y	Y	N	Y
Richter[61]	?	?	Y	N	N	?	Y	Y	Y	?	?
Strender[53]	?	Y	Y	N	?	N	Y	Y	?	N	N

Y: Yes, N: No, ?: unclear because of insufficient information

The study by Smedmark et al[48] fulfilled all criteria for external validity. It showed fair to moderate reliability among two physiotherapists making judgements on stiffness. In the other study that used representative patients, by Pool et al[52] substantial reliability was reached for evaluating motion segment C2-C3 and, overall, reliability was slight to fair.

Two studies fulfilled both criteria for internal validity.[48;67] Strender et al[67] achieved slight reliability for assessing the upper cervical spine in volunteers. The study by Mior et al[56] scored positive on the criterion of stability of characteristics. It showed slight reliability among two pre-trained students of chiropractic examining fixations of vertebra C1 in healthy students. The study by Fjellner et al[44] did not satisfy this criterion because a large number of tests was involved. In their study with healthy volunteers, estimates of reliability ranged from poor to moderate.

Fair to moderate reliability was consistently shown in the one study, by Smedmark et al[48], that was externally and internally valid. The lowest levels of reliability were reached by Christensen et al[50], with values of kappa up to -0.42 for prone joint play evaluation of the upper thoracic spine in non-representative patients. Their estimates could have been biased due to low prevalence.

Assessing mobility of motion segment C1-C2 reached at least a fair level of reliability in five studies.[36;44;47;48;52] Examination of motion segment C2-C3 yielded fair to substantial values of kappa in three studies.[47;48;52]

Lumbar spine (n = 10)
Data on estimates of inter-examiner reliability of PIVM assessment are presented in Table 2 (last column). Inter-examiner reliability for the lumbar spine ranged from poor to substantial. Overall, reliability was poor to fair.

Two studies fulfilled all criteria for external validity.[51;62] Hicks et al[51] showed poor to fair reliability among four pre-trained examiners making judgements on antero-posterior mobility of vertebrae L1-L5 with subjects in prone position. Using this same assessment procedure, Maher & Adams[62] did not find acceptable ICC values.

The study by Hicks et al[51] fulfilled the criteria for internal validity. Systematic error could have biased their estimates. In three studies[53;61;63], stability of characteristics during research was not likely. Richter & Lawall[61] reported a reliability among five physicians ranging from slight to substantial. Strender et al[53] calculated substantial values of weighted kappa for two physiotherapists judging mobility of motion segments L4-L5 and L5-S1 with a side-lying (hips and knees flexed) technique described by Kaltenborn.

Substantial reliability was shown by Strender et al.[53] Their estimates could have been biased due to low prevalence (L4-L5) and systematic error (L5-S1). The lowest levels of reliability, with predominantly negative values of kappa, were reached by Mootz et al[39] for evaluation of fixations in students of chiropractic. Prevalence bias due to limited variation could have influenced their results.

Chiropractic seated motion palpation for intervertebral fixations consistently yielded poor to fair inter-examiner reliability in three studies.[39;41;59]

Discussion

In this systematic review, the inter-examiner reliability of segmental PIVM assessment of the cervical and lumbar spine ranged from poor to substantial. However, overall, reliability was poor to fair.

Studies addressing reliability are conducted to evaluate the consistency of measurements and to quantify measurement error within or between examiners.[5-7;30;70-72] A repeated-measures design consists of one assessment of all subjects by two or more examiners to determine inter-examiner reliability.[73] Inter-examiner reliability reflects a profession's performance.[73] This systematic review was conducted to contribute resolving uncertainty over consistency among manual practitioners in assessing PIVM in the spine.

Only four out of the 19 included studies used representative patients as participants.[48;51;52;62] Estimates of reliability of a test procedure are intimately linked with the population it was used in.[5] In order to assure external validity, it is necessary to include patients with neck or low-back pain that are likely to undergo PIVM assessment procedures in clinical practice.[30] This issue also deals with the essence of the concept of reliability, for reliability can only exist when individuals vary in the characteristic under study like symptomatic subjects most likely do.[5] We note that characteristics of representative patients may differ substantially for the various health care systems depending on the level of direct accessibility of practitioners. The need to use symptomatic participants has also been emphasised by other reviewers.[12;14-16] From evidence of two studies, we found that reliability tended to be higher when representative neck patients were examined.[48;52]

With regard to internal validity, only three studies satisfied both criteria of blinding of examiners to each others' results and stability of joint mobility during research.[48;51;67] Estimates of reliability can only be valid when the characteristic under study does not change during research, otherwise true reliability will be underestimated.[30] Where PIVM assessment is concerned, stability of biomechanical properties of connective tissue during the research process forms a key issue. These properties are susceptible to change as a result of natural variation over time or mobilising effects of the test procedure itself.[30] None of the included studies explicitly dealt with this issue in their design. In the majority of cases, the study protocol was poorly reported. Items such as the number of tests, the number of movement repetitions, forces applied in end-position, motion directions, and time intervals should be considered and described thoroughly. Some researchers post hoc discussed the possibility of changes in mobility as a result of the assessment procedure.[44;53;63;65] One of the excluded studies used a Latin square design to correct for systematic differences in characteristics induced by the test.[42] In their reviews, Huijbregts[15] and Hestbæk & Leboeuf-Yde[16] also recognised the importance of stable characteristics but they did not use this as a quality criterion. In one internally valid study, fair to moderate reliability was consistently shown when representative patients were examined.[48]

Currently, kappa is the statistic of choice for analysing inter-examiner reliability with nominal data.[13;23;27] Most of our included studies appropriately used kappa statistics. However, the interpretation of kappa is not straightforward. Feinstein & Cicchetti described two paradoxes of kappa by examining cross tabulations.[26] The first paradox concerns kappa taking lower values in case of substantial symmetrical imbalance in marginal totals and high percentage agreement. This situation, called limited variation in the presence or absence of a characteristic, makes kappa susceptible for prevalence bias.[6;25] In the second paradox, Kappa overestimates in case of asymmetrical imbalance in marginal totals which is likely when examiners systematically disagree. As a consequence, comparing kappa values from different studies, let alone pooling them, is unjustifiable.[25;29] At least a critical appraisal of possible prevalence bias and systematic bias is required. For this purpose, raw data, like cross tabulations, are indispensable.[6] The majority of studies reviewed did not adequately report statistical data. Appropriate statistical techniques for pooling kappa statistics recognising the problems with prevalence and systematic error are not available. Due to this fact as well as the strong clinical heterogeneity across studies, we did not perform a meta-analysis to summarise reliability.

The concept of prevalence bias is closely related to the choice of study population. This bias is likely when a homogeneous (e.g., asymptomatic) sample is used. In harmony with the need to include representative patients as participants, as stated earlier, careful attention to the choice for a heterogeneous study population will decrease the risk of prevalence bias.[26] Meade et al[75] proposed the Phi (Φ) statistic as a chance-independent statistic to overcome prevalence problems with kappa. In the two studies that showed the lowest levels of reliability, non-representative participants were used and estimates were biased due to low prevalence.[39;50]

With respect to reducing systematic error between examiners, several authors have suggested enhanced standardisation of procedures to reduce error and improve reliability.[5;14;53;67;76] Others have argued that training of examiners diminishes external validity.[15;51] We found no relevant differences in reliability for pre-trained examiners.

Eight studies determined estimates of intra-examiner reliability, but acceptable levels were not reached. We did not conduct a separate appraisal of the internal validity of the intra-examiner reliability designs within our quality assessment. In an intra-examiner reliability design, each examiner performs repeated measurements of each subject. Error within examiners constitutes an integral source of the total amount of error between examiners.[5] Intra-examiner reliability can be computed from an inter-examiner design whilst still avoiding specific problems with blinding, consistency of error, and instability of characteristics under study.[73]

In diagnostic accuracy studies, availability of clinical information from participants to examiners before executing the test has been shown to increase sensitivity.[8] This distortion is known as clinical review bias. In the context of inter-examiner reliability research reflecting daily practice of manual practitioners, this type of bias is likely to occur because the same examiner both gathers clinical information and performs the physical examination. Using the QUADAS tool, examiners are allowed to have clinical information as long as this information reflects clinical practice.[22] In case of analysing reliability with kappa, prior knowledge and expectation may influence calculations.[26;46;56] Furthermore, we argue that not blinding examiners to clinical characteristics will reduce the view on the reliability of the test procedure itself. Therefore, in our quality assessment, we judged the presence of fully blinded examiners as a positive feature.

Seven of the included studies used marking of spinal levels.[39;41;50;57;62;63;65] To date, results of inter-examiner reliability studies on palpating and nominating spinal levels have been inconclusive.[69;77-79] It is unclear whether this pre-conditional skill contributes to another source of error in PIVM assessment.

We consistently found at least fair levels of inter-examiner reliability for PIVM assessment of motion segments C1-C2 and C2-C3, but low values of reliability estimates were found for chiropractic lumbar motion palpation. We could not discover other explanations for heterogeneity in reliability.

Segmental PIVM assessment of the spine is part of the diagnostic clinical expertise of manual practitioners to guide decisions on a therapeutic strategy for patients with neck or low-back pain.[4;62] Hypomobility indicates mobilising interventions while hypermobility calls for a stabilising approach.[51] Clinical rationales rest on segmental approaches.[9] Evidence collected from studies included in this systematic review indicates that the inter-examiner reliability of PIVM assessment of the cervical and lumbar spine is low. However, this review has also exposed some shortcomings of research in this area. Only two studies proved to be externally and internally valid of which one found fair to moderate reliability.[48;51] There is a need for new and valid studies to be conducted. Some evidence suggests that PIVM assessment can be accurate.[80] In a randomised diagnostic trial, on the other hand, Haas et al did not find better outcomes for neck pain patients treated with chiropractic manipulations after segmental end-play assessment.[81] Hence, no final conclusions can be drawn yet regarding the clinical usefulness of PIVM evaluation. Similarly, the contribution of this diagnostic intervention to the effectiveness of manipulative therapies remains unclear.

Limitations of this study

This systematic review has several limitations. In our experience, reliability studies were poorly indexed in databases. The main reason for this may be the inconsistent terminology used in reliability research. In addition, we limited our electronic search for relevant studies to MEDLINE and CINAHL. A quick scan in EMBASE showed only duplicate citation postings. In conclusion, although much effort was put in reference tracing and hand searching, it is not impossible that eligible studies were missed. Furthermore, unpublished studies were not included. Publication bias can form a real threat to the internal validity of systematic reviews of reliability studies.

Quality assessment was performed by using a criteria list mainly derived from the assessment of diagnostic accuracy studies. No evidence is available on whether these items also apply in the context of reliability. Empirical evidence of bias, especially concerning blinding of examiners and stability of characteristics during research, is lacking.

Finally, assigning value labels for ranges of kappa was done in accordance with Landis & Koch.[33] As stated by these authors, this classification is an arbitrary one. Others have questioned its appropriateness.[6;74] Using another classification may have yielded different results.

Conclusions and recommendations

In this systematic review, it was found that the inter-examiner reliability of segmental PIVM assessment of the cervical and lumbar spine by manual practitioners was low. However, most studies did not fulfil the criteria for external and internal validity. In general, reporting of study protocol and statistical data was inadequate. In addition, only a few of all possible assessment techniques have been investigated so far. We propose the following recommendations for future research:

- Include representative neck or low-back pain patients as participants that are likely to undergo the assessment procedure in clinical practice, instead of students, volunteers, healthy individuals, or samples with a mix of symptomatic and non-symptomatic subjects;
- Give careful consideration ensuring stability of joint mobility during the research;
- Determine intra-examiner reliability along in the process;
- Present cross tabulations when using kappa statistics to allow for appraisal of prevalence bias and systematic bias;
- Report the study by following the STARD statement.[20;21]

Only when new and valid evidence emerges, uncertainty over diagnostic performance can be resolved and more definitive conclusions can be drawn regarding the clinical usefulness of PIVM assessment. Until then, questions remain about professional credibility and accountability of this diagnostic procedure within evidence-based clinical decision-making in manual therapy.

References

1. Nachemson A, Waddell G, Norlund AI. Epidemiology of neck and low back pain. In: Nachemson A, Jonsson E (editors). Neck and back pain: the scientific evidence of causes, diagnosis, and treatment. Philadelphia, PA: Lippincot, Williams and Wilkins, 2000.
2. Faas A, Chavannes AW, Koes BW, Van den Hoogen JMM, Mens JMA, Smeele LJM, Romeijnders ACM, Van der Laan JR. Dutch College of General Practitioners Standard "low back pain". Huisarts Wet 1996;39:18-31. [In Dutch]
3. Borghouts AJ, Janssen HJ, Koes BW, Muris JWM, Metsemakers JFM, Bouter LM. The management of chronic pain in general practice: a retrospective study. Scand J Prim Health Care 1999;17:215-20.
4. Jull G, Treleaven J, Versace G. Manual examination: is pain provocation a major cue for spinal dysfunction? Aust J Physiother 1994;40:159-65.
5. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use (3rd edition). Oxford, UK: Oxford University Press, 2003.
6. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304:1491-4.
7. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
8. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. A systematic review. Ann Intern Med 2004;140:189-202.
9. Breen A. The reliability of palpation and other diagnostic methods. J Manipulative Physiol Ther 1992;15:54-6.
10. Maher C, Latimer J. Pain or resistance – the manual therapists' dilemma. Aust J Physiother 1992;38:257-60.
11. Troyanovich SJ, Harrison DD. Motion palpation: it's time to accept the evidence. J Manipulative Physiol Ther 1998;21:568-71.
12. Keating JC. Interexaminer reliability of motion palpation of the lumbar spine: a review of quantitative literature. Am J Chiropr Med 1989;2:107-10.
13. Haas M. The reliability of reliability. J Manipulative Physiol Ther 1991;14:199-208.

14. Panzer DM. The reliability of lumbar motion palpation. J Manipulative Physiol Ther 1992;15:518-24.
15. Huijbregts PA. Spinal motion palpation: a review of reliability studies. J Man Manipulative Ther 2002;10:24-39.
16. Hestbæk L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther 2000;23:258-75.
17. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain. A systematic review of the literature. Spine 2004; 29:E413-25.
18. Egger M, Davey Smith G, O'Rourke K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Davey Smith G, Altman DG (editors). Systematic reviews in health care: meta-analysis in context (2nd edition). London, UK: BMJ Publishing Group, 2001.
19. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6.
20. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. Clin Chem 2003;49:1-6.
21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
22. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology 2003;3:25.
23. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20:37-46.
24. Maclure M, Willett WC. Misinterpretation and misuse of the Kappa statistic. Am J Epidemiol 1987;126:161-9.
25. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. J Clin Epidemiol 1988;41:949-58.
26. Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.

27. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991.
28. Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991;14:119-32.
29. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993;46:423-9.
30. Rothstein JM, Echternach JL. Primer on measurement: an introductory guide to measurement issues. Alexandria, VA: American Physical Therapy Association, 1993.
31. Potter NA, Rothstein JM. Intertester reliability for selected clinical tests of the sacroiliac joint. Phys Ther 1985;65:1671-5.
32. Meijne W, Van Neerbos K, Aufdemkampe G, Van der Wurff P. Intraexaminer and interexaminer reliability of the Gillet test. J Manipulative Physiol Ther 1999;22:4-9.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
34. Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111-23.
35. Johnston WL, Blum GA, Hendra JL, Neff DR, Rosen ME. Passive gross motion testing: Part III. Examiner agreement on selected subjects. J Am Osteopath Assoc 1982;81:309-13.
36. DeBoer KF, Harmon R, Tuttle CD, Wallace H. Reliability study of detection of somatic dysfunction in the cervical spine. J Manipulative Physiol Ther 1985;8:9-16.
37. Love RM, Brodeur RR. Inter- and intra-examiner reliability of motion palpation for the thoracolumbar spine. J Manipulative Physiol Ther 1987;10:1-4.
38. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 1987;67:1526-32.
39. Mootz RD, Keating JC, Kontz HP, Milus TB, Jacobs GE. Intra- and interobserver reliability of passive motion palpation of the lumbar spine. J Manipulative Physiol Ther 1989;12:440-45.
40. Nansel DD, Peneff AC, Jansen RD, Cooperstein R. Interexaminer concordance in detecting joint-play asymmetries in the cervical spines of otherwise asymptomatic subjects. J Manipulative Physiol Ther 1989;12:428-33.
41. Keating JC, Bergmann TF, Jacobs GE, Finer BA, Larson K. Interexaminer reliability of eight evaluative dimensions of lumbar segmental abnormality. J Manipulative Physiol Ther 1990;13:463-70.

42. Binkley J, Stratford PW, Gill C. Interrater reliability of lumbar accessory motion mobility testing. Phys Ther 1995;75:786-95.
43. Van Dillen LR, Sahrmann SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK, Woolsey NB. Reliability of physical examination items used for classification of patients with low back pain. Phys Ther 1998;78:979-88.
44. Fjellner A, Bexander C, Faleij R, Strender L-E. Interexaminer reliability in physical examination of the cervical spine. J Manipulative Physiol Ther 1999;22:511-6.
45. Hawk C, Phongphua C, Bleecker J, Swank L, Lopez D, Rubley T. Preliminary study of the reliability of assessment procedures for indications for chiropractic adjustments of the lumbar spine. J Manipulative Physiol Ther 1999;22:382-9.
46. French SD, Green S, Forbes A. Reliability of chiropractic methods commonly used to detect manipulable lesions in patients with chronic low-back pain. J Manipulative Physiol Ther 2000;23:231-8.
47. Schöps P, Pfingsten M, Siebert U. Reliabilität manualmedizinischer Untersuchungstechniken an der Halswirbelsäule. Studie zur Qualitätssicherung in der manuellen Diagnostik. Z Orthop Ihre Grenzgeb 2000;138:2-7.
48. Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. Man Ther 2000;5:97-101.
49. Van Suijlekom HA, de Vet HCW, van den Berg SGM, Weber WEJ. Interobserver reliability in physical examination of the cervical spine in patients with headache. Headache 2000;40:581-6.
50. Christensen HW, Vach W, Vach K, Manniche C, Haghfelt T, Hartvigsen L, Høilund-Carlsen PF. Palpation of the upper thoracic spine: an observer reliability study. J Manipulative Physiol Ther 2002;25:285-92.
51. Hicks GE, Fritz JM, Delitto A, Mishock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. Arch Phys Med Rehabil 2003;84:1858-64.
52. Pool JJ, Hoving JL, de Vet HC, van Mameren H, Bouter LM. The interexaminer reproducibility of physical examination of the cervical spine. J Manipulative Physiol Ther 2004;27:84-90.
53. Strender L-E, Sjöblom A, Sundell K, Ludwig R, Taube A. Interexaminer reliability in physical examination of patients with low back pain. Spine 1997;22:814-20.

54. Kaltenborn F, Lindahl O. Reproducerbarheten vid rörelseundersökning av enskilda kotor. Lakartidningen 1969;66:962-5.
55. Gonella C, Paris SV, Kutner M. Reliability in evaluating passive intervertebral motion. Phys Ther 1982;62:436-44.
56. Mior SA, King RS, McGregor M, Bernard M. Intra and interexaminer reliability of motion palpation in the cervical spine. J Can Chiropractic Association 1985;29:195-8.
57. Bergström E, Courtis G. An inter- and intra-examiner reliability study of motion palpation of the lumbar spine in lateral flexion in the seated position. Eur J Chiropractic 1986;34:121-41.
58. Jull G, Bullock M. A motion profile of the lumbar spine in an ageing population assessed by manual examination. Physiotherapy Practice 1987;3:70-81.
59. Boline PD, Keating JC, Brist J, Denver G. Interexaminer reliability of palpatory evaluations of the lumbar spine. Am J Chiropr Med 1988;1:5-11.
60. Leboeuf C, Gardner V, Carter AL, Scott TA. Chiropractic examination procedures: a reliability and consistency study. J Australian Chir Assoc 1989;19:101-4.
61. Richter T, Lawall J. Zur Zuverlässigkeit manualdiagnostischer Befunde. Manuelle Med 1993;31:1-11.
62. Maher C, Adams R. Reliability of pain and stiffness assessments in clinical manual lumbar spine examination. Phys Ther 1994;74:801-11.
63. Inscoe EL, Witt PL, Gross MT, Mitchell RU. Reliability in evaluating passive intervertebral motion of the lumbar spine. J Man Manip Ther 1995;3:135-43.
64. Schoensee SK, Jensen G, Nicholson G, Gossman M, Katholi C. The effect of mobilization on cervical headaches. J Orthop Sports Phys Ther 1995;21:184-96.
65. Haas M, Raphael R, Panzer D, Peterson D. Reliability of manual end-play palpation of the thoracic spine. Chiropr Tech 1995;7:120-4.
66. Jull G, Zito G, Trott P, Potter H, Shirley D, Richardson C. Inter-examiner reliability to detect painful upper cervical joint dysfunction. Aust J Physiother 1997;43:125-9.
67. Strender L-E, Lundin M, Nell K. Interexaminer reliability in physical examination of the neck. J Manipulative Physiol Ther 1997;20:516-20.
68. Hanten WP, Olson SL, Ludwig GM. Reliability of manual mobility testing of the upper cervical spine in subjects with cervicogenic headache. J Man Manip Ther 2002;10:76-82.

69. Downey B, Taylor N, Niere K. Can manipulative physiotherapists agree on which lumbar level to treat based on palpation? Physiotherapy 2003;89:74-81.
70. Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ 1996;313:41-2.
71. Bland JM, Altman DG. Measurement error. BMJ 1996;313:744.
72. Bruton A, Conway JH, Holgate ST. Reliability: what is it and how is it measured? Physiotherapy 2000;86:94-9.
73. Haas M. How to evaluate intraexaminer reliability using an interexaminer reliability study design. J Manipulative Physiol Ther 1995;18:10-5.
74. Lantz CA. Application and evaluation of the Kappa statistic in the design of chiropractic clinical research. J Manipulative Physiol Ther 1997;20:521-8.
75. Meade MO, Cook RJ, Guyatt GH, Groll R, Kachura JR, Bedard M, Cook DJ, Slutsky AS, Stewart TE. Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. Am Journal Resp Crit Care Med 2000;161:85-90.
76. Anson E, Cook C, Camacho C, Gwilliam B, Karakostas T. The use of an educational model in the improvement of student reliability in finding R1. J Man Manip Ther 2003;11:204-12.
77. McKenzie AM, Taylor NF. Can physiotherapists locate lumbar spinal levels by palpation? Physiotherapy 1997;83:235-9.
78. Downey BJ, Taylor NF, Niere KR. Manipulative physiotherapists can reliably palpate nominated lumbar spinal levels. Man Ther 1999;4:151-6.
79. Billis EV, Foster NE, Wright CC. Reproducibility and repeatability: errors of three groups of physiotherapists in locating spinal levels by palpation. Man Ther 2003;8:223-32.
80. Humphreys BK, Delahaye M, Peterson CK. An investigation into the validity of cervical spine motion palpation using subjects with congenital block vertebrae as a 'gold standard'. BMC Musculoskelet Disord 2004;5:19.
81. Haas M, Groupp E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay assessment as an indicator for spinal manipulation. Spine 2003;28:1091-6.

Chapter 3

Inter-examiner reliability of passive motion assessment of the extremities

Chapter 3a

Inter-examiner reliability for measurement of passive physiological range of motion in upper extremity joints is better if instruments are used: A systematic review

Rachel van de Pol, Emiel van Trijffel, Cees Lucas

J Physiother 2010;56:7-17

Abstract

Background: Passive assessment of motion in joints of the upper extremity is commonly used by physiotherapists in order to measure joint restrictions and to diagnose musculoskeletal disorders. To date, no systematic appraisal of studies on the inter-examiner reliability of measurement of passive movements in upper extremity joints has been conducted.

Methods: We conducted a systematic review of studies of inter-examiner reliability of measurements of passive movement in upper extremity joints published up to July 1, 2009. Studies involving participants with and without upper extremity disorders were included. Range of motion measurements and end-feel judgements from passive joint motion examination using methods and instruments feasible in clinical practice were considered. No language restrictions were imposed. Study selection, quality assessment, and data extraction were performed by two reviewers independently.

Results: Twenty-one studies were included of which 11 demonstrated acceptable inter-examiner reliability. Two studies satisfied all criteria for internal validity while reporting almost perfect reliability. Overall, the methodological quality of studies was poor. ICC ranged from 0.26 (95% CI -0.01 to 0.69), for measuring the physiological range of shoulder internal rotation using vision, to 0.99 (95% CI 0.98 to 1.0), for the physiological range of finger and thumb flexion/extension using a goniometer. Measurements of physiological range of motion using instruments were more reliable than using vision. Measurements of physiological range of motion were also more reliable than judgements of end-feel or of accessory range of motion.

Conclusions: Inter-examiner reliability for the measurement of passive movements of upper extremity joints varies with the method of measurement. In order to make reliable decisions about joint restrictions in clinical practice, we recommend that clinicians measure passive physiological range of motion using goniometers or inclinometers.

Introduction

Physiotherapists commonly assess and treat upper extremity disorders. Passive joint mobilisation or manipulation has been shown to be effective in disorders such as adhesive shoulder capsulitis, non-specific shoulder pain or dysfunction, shoulder impingement syndrome, lateral epicondylalgia, and carpal tunnel syndrome.[1-4] Measurement of passive movement is indicated in order to assess joint restrictions and to help diagnose these disorders. Passive movement, either physiological or accessory, can be reported as range of motion, end-feel, or pain and is an indication of the integrity of joint structures.[5;6] Passive physiological range of motion may be measured using vision or instruments such as goniometers or inclinometers.

An essential requirement of clinical measures is that they are valid and reliable so that they can be used to discriminate between individuals.[7] Inter-examiner reliability is a component of reproducibility along with agreement and refers to the relative measurement error, i.e., the variation between patients as measured by different examiners in relation to the total variance of the measurements.[7] Agreement, on the other hand, provides insight into the ability of a clinical measure to yield the same value on multiple occasions and reflects the absolute measurement error.[8] High inter-examiner reliability for measurements of upper extremity joints is a prerequisite for valid and uniform decisions about joint restrictions.[9]

Many studies investigating the reliability of passive movements of human joints have been conducted. However, relatively few reviews have summarised and appraised the evidence. For example, seven systematic reviews have been published on passive spinal movement.[10-16] In general, inter-examiner reliability was low and studies were of poor methodological quality. To date, no systematic appraisal of studies on inter-examiner reliability of measurements of passive movement in upper extremity joints has been conducted. Therefore, the research question for this systematic review was: What is the inter-examiner reliability for measurements of passive physiological or accessory movements in upper extremity joints?

Methods

Study selection

MEDLINE (PubMed) was searched by two reviewers (RvdP, EvT) independently for studies published between January 1, 1966, and July 1, 2009. Search terms included all relevant upper extremity joints and all synonyms for *reliability* and *examiner* (Appendix 1). Additional searches in CINAHL (1982 to July 1, 2009) and EMBASE (1996 to July 1, 2009) were performed by one reviewer (RvdP). In addition, reference lists of all retrieved papers were hand searched for relevant studies. Additionally retrieved studies were checked for eligibility by the second reviewer.

The titles and abstracts were screened by two reviewers (RvdP, EvT) independently. When relevant, full text papers were retrieved. Studies were included if they met all inclusion criteria (Box 1).

Box 1. Inclusion criteria

Design
- Repeated measures between examiners

Participants
- Symptomatic and asymptomatic individuals

Measurement procedure
- Performed passive (i.e., manual) physiological or accessory movements in any of the joints of the shoulder, elbow, or wrist-hand-fingers
- Reported range of motion or end-feel
- Used methods feasible in clinical practice (considering instruments, costs, amount of training required)

Outcomes
- Estimates of inter-examiner reliability

No restrictions were imposed on language or date of publication. Abstracts and documents that were anecdotal, speculative, or editorial in nature were not included. Studies investigating active movement or restriction in passive motion due to pain or ligament stability as well as animal or cadaver studies were not considered for inclusion. Studies of people with neurological conditions in which abnormal muscle tone may interfere with joint movement, or of people after arthroplasty were also excluded. Disagreements on eligibility were first resolved by discussion and decided by a third reviewer (CL) if disagreement persisted.

Quality assessment

No validated instrument was available for assessing methodological quality of inter-examiner reliability studies. Therefore, a list of criteria for quality was compiled from the QUADAS tool, the STARD Statement, and criteria used for assessing studies on reliability of passive spinal movements.[15;17-19] Criteria were rated as 'Yes', 'No', or '?' (unclear because of insufficient information) (Box 2).

Box 2. Criteria for assessing methodological quality

1. Was a representative sample of participants used?
2. Was a representative sample of examiners used?
3. Is replication of the measurement procedure possible?
4. Was clinical information from participants available to examiners and comparable to clinical practice?
5. Were participants' characteristics stable during the study?
6. Were examiners' characteristics stable during the study?
7. Were examiners blinded to each other's results?
8. Can non-random loss to follow-up be ruled out?
9. Was an estimate of intra-examiner reliability validly determined and was it above 0.80?
10. Were appropriate measures (kappa, intraclass correlation coefficient) used for calculating reliability?

Criteria 1 to 4 assess external validity, Criteria 5 to 9 assess internal validity, and Criterion 10 assesses statistical methods. External validity was considered sufficient if Criteria 1 to 4 were rated 'Yes'. With respect to internal validity, Criteria 5 to 7 were assumed to be decisive in determining risk of bias. A study was considered to have a low risk of bias if Criteria 5 to 7 were all rated 'Yes', a moderate risk if two of these criteria were rated 'Yes', and a high risk if none or only one of these criteria were rated 'Yes'. After training, two reviewers (RvdP, EvT) independently assessed methodological quality of all included studies and were not blind to journal, authors, and results. If discrepancy between reviewers persisted after discussion, a decisive judgement was passed by the third reviewer (CL).

Data extraction

We extracted data on participants (number, age, clinical characteristics), examiners (number, profession, training), measurements (joints and movement direction, movement performed, method, outcomes reported), and inter-examiner reliability (point estimates, estimates of precision). Two reviewers (RvdP, EvT) extracted data independently and were not blind to journal, authors, or results. When disagreement between reviewers could not be resolved by discussion, a third reviewer (CL) made the final decision.

Data analysis

Data were analysed by examining intraclass correlation coefficient (ICC) and Cohen's kappa (95% CI). ICC >0.75 indicates an acceptable level of reliability.[20] Corresponding kappa levels were used as assigned by Landis & Koch where <0.00 = poor, 0.00-0.20 = slight, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = substantial, and 0.81-1.00 = almost perfect.[21] In addition, reliability was analysed relating it to methodological quality and risk of bias. Reliability from studies not fulfilling Criteria 5 or 6 could have been underestimated while reliability from studies not fulfilling Criterion 7 could have been overestimated. Negative scores on combinations of Criteria 5-7 could have led to bias in an unknown direction. Where one or more of these three criteria were unclear, no statement was made regarding the presence or direction of potential bias. Finally, because of clinical and methodological heterogeneity between studies, we did not attempt to statistically summarise data by calculating pooled estimates of reliability.

Results

Flow of studies through the review

Searching MEDLINE yielded 326 citations of which 26 papers were retrieved in full text. CINAHL (95 citations) and EMBASE (34) yielded no additional relevant articles. Hand searching supplied another 20 potentially relevant studies. Of these 46, 25 studies were excluded (Appendix 2). In total, 21 studies fulfilled all inclusion criteria (Figure).

Characteristics of included studies

The included studies are summarised in Table 1. Thirteen studies[22-34] investigated inter-examiner reliability of measurement of shoulder movements, two[35;36] investigated elbow movements, four[37-40] investigated wrist movements, one[41] investigated phalangeal joint movements, and one[42] investigated thumb movements. In all except two studies[37;42], physiotherapists acted as examiners. There were no disagreements between reviewers on selection of studies.

Quality of studies

The methodological quality of included studies is presented in Table 2.

One study[29] fulfilled all four criteria for external validity and four studies satisfied three criteria. Two studies[30;41] fulfilled all three criteria for internal validity representing a low risk of bias while six studies satisfied two criteria. Criteria on internal and external validity could not be scored on 54/189 (29%) occasions because of insufficient reporting. Twenty/210 (10%) disagreements occurred between reviewers which were all resolved by discussion.

Inter-examiner reliability by region

The inter-examiner reliability for measurements of passive physiological range of motion is presented in Table 3, and for judgements of accessory range of motion and physiological end-feel in Tables 4 and 5, respectively.

Shoulder (n = 13)

One study[29] fulfilled all criteria for external validity and another[30] fulfilled all criteria for internal validity. ICC for measurement of physiological range of motion using vision ranged from 0.26 (95% CI -0.01 to 0.69) for internal rotation[25] to 0.96 for abduction[30]. In seven studies[23;24;27-30;33], acceptable reliability (ICC >0.75) was reached. The highest reliability

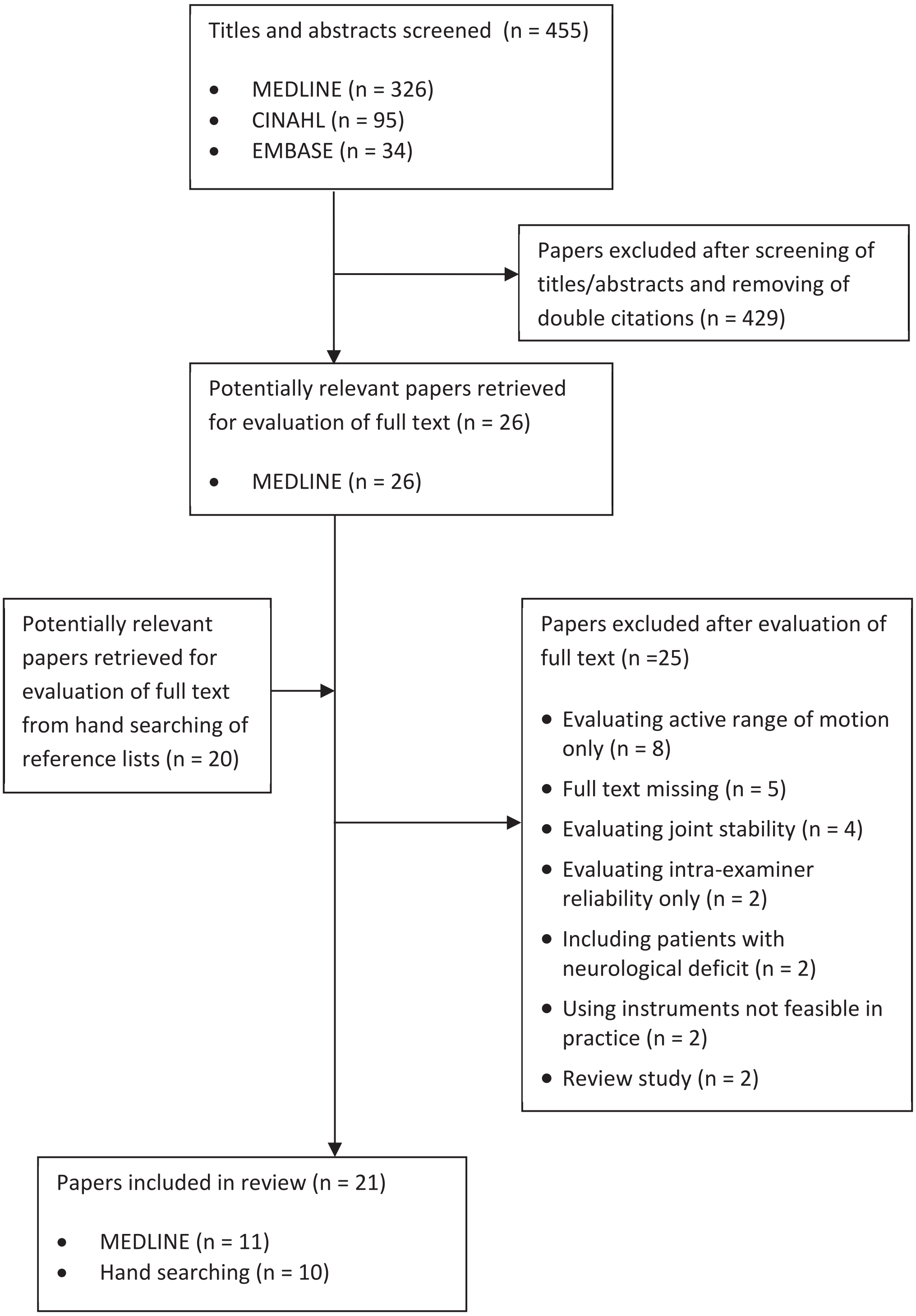

Figure. Flow of studies through the review

occurred in Nomden et al[30] for experienced physiotherapists, of which one was a specialist in manual therapy, in patients with shoulder pathology and it was associated with a low risk of bias. In general, measuring passive physiological range of motion using instruments

Table 1. Characteristics of included studies *(n = 21)*

Study	Participants	Examiners	Joints Movement direction	Position	Movement performed	Method	Outcome reported	Reliability statistic
Awan[22]	n = 56 Age = range 13-18 yr Condition = normal	n = 4 Profession = 2 physiatrists, 1 PT, 1 resident doctor Training = Y	Shoulder • IR • ER	 Supine Sh 90 deg Abd	Physiological	Digital inclinometer Vision	ROM	ICC
Bovens[37]	n = 148 Age = mean 48 yr (SD 7) Condition = normal	n = 3 Profession = physician Training = Y	Wrist-hand-fingers • Wrist F • Wrist E	 Palms together Hands together	Physiological	Vision	ROM	R
Chesworth[23]	n = 34 Age = mean 55 yr (SD 18.5) Condition = shoulder pathology, post-surgery	n = 2 Profession = PT/MT Training = N	Shoulder • ER	 Supine Sh 20° Abd Elbow 90° F	Physiological	Vision Manual	ROM End-feel	ICC (2, 1)
De Kraker[42]	n = 25 Age = mean 30 yr (SD 7) Condition = normal	n = 2 Profession = 1 HT, 1 trainee plastic and reconstructive surgery Training = N	Wrist-hand-fingers • Thumb Abd	 Seated Elbow 90° FL Wrist neutral	Physiological	Pollexograph Goniometer	ROM	ICC
De Winter[24]	n = 155 Age = mean 47 yr (SD 12.6) Condition = shoulder pathology	n = 2 Profession = PT Training = Y	Shoulder • Abd • ER	 Seated Supine	Physiological	Digital inclinometer	ROM	ICC
Glasgow[41]	n = 10 Age = mean 39.7 yr (SD 13.5) Condition = traumatic hand injuries	n = 2 Profession = ? Training = N	Hand-wrist-fingers • IP F • IP E • MCP F	?	Physiological	Goniometer	ROM	ICC (2,1)

			• MCP E • Thumb F • Thumb E					
Hayes[25]	n = 8 Age = mean 66 yr (SD 5.7) Condition = shoulder pathology, post-surgery	n = 4 Profession = 2 PT, 1 orthopaedic surgeon, 1 sports physician Training = Y	Shoulder • F • Abd • ER • IR	Seated	Physiological	Vision	ROM	ICC (2, 1)
Hayes[26]	n = 18 Age = mean 34.3 yr (SD 12.9) Condition = shoulder pain	n = 2 Profession = PT Training = Y	Shoulder • Abd • ER • IR • Hor Add • Full Abd	Standing (with and without scapular stabilization)	Physiological	Manual	End-feel	kappa
Heemskerk[27]	n = 12 Age = mean 36 yr (range 25-49) Condition = normal	n = 2 Profession = PT Training = N	Shoulder • Abd • ER	Seated Supine	Physiological	Digital inclinometer	ROM	ICC
Horger[38]	n = 48 Age = mean 38.8 yr (range 18-71) Condition = wrist injuries	n = 26 Profession = 11 OT, 2 PT, 6 HT, 7 non- specialised examiners Training = N	Wrist • E • F • Abd • Add	?	Physiological	Goniometer	ROM	ICC (1,1)
LaStayo[39]	n = 140 Age = mean 41.5 yr (range 6-81) Condition = wrist pathology	n = 32 Profession = 25 OT, 6 PT, 1 OT/PT (17 of which HT) Training = N	Wrist • E • F	?	Physiological	Goniometer	ROM	ICC (2,1)
Lin[28]	n = 16 Age = mean 54.5 yr (SD 9.2) Condition = shoulder stiffness	n = 2 Profession = PT Training = N	Shoulder • Hor F • Hor E	Supine Sh 90° F Sh 0° Add Scapula stabilised	Physiological	Inclinometer	ROM	ICC (3,1)

MacDermid[29]	n = 34 Age = mean 55 yr (SD 18) Condition = shoulder pathology, post-surgery	n = 2 Profession = PT/MT Training = N	Shoulder • ER	Supine Sh 20° to 30° Abd Elbow 90° F	Physiological	Goniometer	ROM	ICC
Nomden[30]	n = 91 Age = mean 48.5 yr (SD 11.8) Condition = shoulder pathology	n = 2 Profession = 1 PT, 1 PT/MT Training = N	Shoulder • Abd • ER	Seated Abd: Sh 0° Abd Sh ER Thumb up ER: Sh 0° F Elbow 90° F	Physiological	Vision	ROM	ICC (1,1)
Patla[35]	n = 20 Age = ? Condition = normal, elbow pathology	n = 2 Profession = PT Training = Y	Elbow • E • F	Standing Sh 20° Abd Elbow 20° F	Physiological	Goniometer Manual	ROM End-feel	kappa
Riddle[31]	n = 50 Age = mean 48.6 yr (SD 14.4) Condition = shoulder pathology	n = 16 Profession = PT Training = N	Shoulder • F • E • Abd • Hor Abd • Hor Add • ER • IR	Supine Prone Seated Side lying Standing	Physiological	Goniometer	ROM	ICC (1,1)
Rothstein[36]	n = 12 Age = ? Condition = elbow pathology	n = 12 Profession = PT Training = N	Elbow • E • F	?	Physiological	Goniometer	ROM	ICC
Staes[40]	n = 30, 15 Age = mean 21.3 yr (SD 1.6), mean age 38.3 yr (SD 11) Condition = normal, wrist pathology	n = 2 Profession = PT Training = Y	Wrist • Hamate • Lunate • Scaphoid • Trapezoid against Capitate	Resting position	Accessory	Vision	ROM End-feel	Weighted kappa

Terwee[32]	n = 201 Age = mean 48yr (SD 12) Condition = shoulder pathology	n = 2 Profession = PT Training = Y	Shoulder • ELE • Abd • ER • Hor Add	Seated Sh 0° F Elbow 90° F	Physiological	Vision	ROM	ICC (2,1)
Tyler[33]	n = 28 Age = mean 30 yr (SD 8.9) Condition = normal	n = 2 Profession = PT Training = N	Shoulder • Hor F	Side lying Sh 90° Abd Scapula stabilised	Physiological	Measuring tape	ROM	ICC (3,k)
Van Duijn[34]	n = 18 Age = mean 36.6 yr (SD 10) Condition = shoulder pathology, normal	n = 6 Profession = PT Training = N	Shoulder • Inferior glide	Supine	Accessory	Vision	ROM	ICC (2,1)

Abd: abduction, Add: adduction, ELE: elevation, ER: external rotation, E: extension, F: flexion, HT: hand therapist, Hor: horizontal, IP: interphalangeal, ICC: intraclass correlation coefficient, IR: internal rotation, MCP: metacarpophalangeal, MT: manual therapist, N: No, OT: occupational therapist, PT: physiotherapist, ROM: range of motion, Sh: shoulder, Y: Yes

Table 2. Quality of studies *(n = 21)*

	External validity				Internal validity					Statistical methods
Study	1	2	3	4	5	6	7	8	9	10
Awan[22]	N	?	Y	N	Y	?	Y	Y	N	?
Bovens[37]	N	?	Y	?	Y	Y	?	Y	?	?
Chesworth[23]	Y	N	Y	Y	N	?	Y	Y	N	Y
De Kraker[42]	N	?	Y	?	?	N	Y	Y	N	?
De Winter[24]	Y	?	Y	?	Y	Y	?	N	N	?
Glasgow[41]	Y	?	Y	N	Y	Y	Y	N	Y	Y
Hayes[25]	Y	?	Y	?	N	?	Y	Y	N	Y
Hayes[26]	N	Y	Y	Y	?	?	Y	Y	N	?
Heemskerk[27]	N	?	N	N	?	N	Y	Y	N	?
Horger[38]	Y	Y	N	?	?	?	Y	Y	N	Y
LaStayo[39]	Y	Y	Y	N	?	?	Y	Y	N	Y
Lin[28]	Y	?	Y	?	?	?	Y	Y	N	Y
MacDermid[29]	Y	Y	Y	Y	?	?	Y	Y	N	?
Nomden[30]	Y	Y	Y	?	Y	Y	Y	Y	?	Y
Patla[35]	N	?	Y	?	Y	?	?	Y	?	Y
Riddle[31]	Y	Y	?	N	?	?	Y	Y	N	Y
Rothstein[36]	?	Y	?	N	?	?	Y	Y	N	?
Staes[40]	N	?	Y	N	?	Y	Y	N	N	?
Terwee[32]	Y	Y	Y	N	Y	?	Y	Y	?	Y
Tyler[33]	N	?	Y	N	Y	?	?	Y	Y	Y
Van Duijn[34]	N	Y	Y	N	Y	?	Y	Y	N	N

Y: Yes, N: No, ?: unclear because of insufficient information

such as goniometers or inclinometers resulted in higher reliability than using vision. Of the four studies[22;24;32;34] classified as having a moderate risk of bias, one[24] reported acceptable reliability for measuring abduction (ICC 0.83) and external rotation (0.90). The externally valid study by MacDermid et al[29] reported acceptable reliability (ICC 0.86, 95% CI 0.72 to

Table 3. Inter-examiner reliability (95% CI) for measurement of passive physiological range of motion by method of measurement, joint, and movement direction

Method of measurement	Study	Inter-examiner reliability
Inclinometer		
Shoulder		
External rotation	Awan[22]	ICC 0.41, 0.51
	De Winter[24]	ICC 0.90
	Heemskerk[27]	ICC 0.81 to 0.87
Internal rotation	Awan[22]	ICC 0.50 to 0.66
Abduction	De Winter[24]	ICC 0.83
	Heemskerk[27]	ICC 0.27 to 0.84
Horizontal flexion	Lin[28]	ICC 0.82 (0.54 to 0.94)
Horizontal extension	Lin[28]	ICC 0.89 (0.69 to 0.96)
Goniometer		
Shoulder		
External rotation	MacDermid[29]	ICC 0.85 (0.73 to 0.91), 0.86 (0.72 to 0.92)
	Riddle[31]	ICC 0.88, 0.90
Internal rotation	Riddle[31]	ICC 0.53, 0.55
Abduction	Riddle[31]	ICC 0.84, 0.87
Horizontal abduction	Riddle[31]	ICC 0.28, 0.30
Horizontal adduction	Riddle[31]	ICC 0.35, 0.41
Flexion	Riddle[31]	ICC 0.87, 0.89
Extension	Riddle[31]	ICC 0.26, 0.27
Elbow		
Flexion	Rothstein[36]	ICC 0.85 to 0.97
Extension	Rothstein[36]	ICC 0.92 to 0.95
Wrist-hand-fingers		
Wrist flexion	Horger[38]	ICC 0.86 (0.78 lower limit)
	LaStayo[39]	ICC 0.88 to 0.93
Wrist extension	Horger[38]	ICC 0.84 (0.75 lower limit)
	LaStayo[39]	ICC 0.80 to 0.84
Wrist abduction	Horger[38]	ICC 0.66 (0.51 lower limit)
Wrist adduction	Horger[38]	ICC 0.83 (0.74 lower limit)

Thumb abduction	De Kraker[42]	ICC 0.37 (-0.42 to 0.79)
Finger/thumb flexion and extension	Glasgow[41]	ICC 0.99 (0.98 to 1.0)
Vision		
Shoulder		
External rotation	Chesworth[23]	ICC 0.83 (0.70 to 0.90), 0.90 (0.83 to 0.95)
	Hayes[25]	ICC 0.57 (0.26 to 0.87)
	Nomden[30]	ICC 0.70
	Terwee[32]	ICC 0.73 (0.22 to 0.88)
Internal rotation	Awan[22]	ICC 0.51, 0.65
	Hayes[25]	ICC 0.26 (-0.01 to 0.69)
Abduction	Hayes[25]	ICC 0.66 (0.37 to 0.90)
	Nomden[30]	ICC 0.96
	Terwee[32]	ICC 0.67 (0.35 to 0.81)
Horizontal adduction	Terwee[32]	ICC 0.36 (0.22 to 0.48)
Flexion	Hayes[25]	ICC 0.70 (0.42 to 0.92)
Elevation	Terwee[32]	ICC 0.87 (0.83 to 0.90)
Wrist-hand fingers		
Wrist flexion	Bovens[37]	R 0.59
Wrist extension	Bovens[37]	R 0.09
Tape measure		
Shoulder		
External rotation	Tyler[33]	ICC 0.80
Pollexograph		
Wrist-hand-fingers		
Thumb abduction	De Kraker[42]	ICC 0.59 (0.42 to 0.89)

ICC: intraclass correlation coefficient, R: correlation coefficient

0.92 and 0.85, 95% CI 0.73 to 0.91) for two experienced physiotherapists with advanced manual therapy training measuring external rotation in symptomatic individuals. In the one study[34] investigating accessory range of motion of the glenohumeral joint (inferior gliding), reliability was found to be unacceptable (ICC 0.52). Overall, measurements of range of motion were more reliable than judgements of end-feel. Kappa for end-feel ranged from

Table 4. Inter-examiner reliability (95% CI) for judgements of passive accessory range of motion by joint and movement direction

Accessory motion	Study	Inter-examiner reliability
Shoulder		
Inferior glide	Van Duijn[34]	ICC 0.52
Wrist-hand-fingers		
Wrist capitate	Staes[40]	κ_w 0.29 to 0.42, 0.33 to 0.87

ICC: intraclass correlation coefficient, κ_w: weighted kappa

0.26 (95% CI -0.16 to 0.68) in full shoulder abduction to 0.70 (95% CI 0.31 to 1.0) in abduction with scapula stabilisation.[26] No specific movement direction was consistently associated with high or low reliability.

Table 5. Inter-examiner reliability (95% CI) of judgements of physiological end-feel by joint and movement direction

Method of assessment	Study	Inter-examiner reliability
Shoulder		
External rotation	Chesworth[23]	ICC 0.34 (0.05 to 0.57) to 0.91 (0.84 to 0.95)
	Hayes[26]	κ 0.47 (0.08 to 0.87)
Internal rotation	Hayes[26]	κ 0.41 (0.03 to 0.80)
Abduction	Hayes[26]	κ 0.70 (0.31 to 1.0)
Horizontal adduction	Hayes[26]	κ 0.40 (0.01 to 0.79)
Full abduction	Hayes[26]	κ 0.26 (-0.16 to 0.68)
Elbow		
Flexion	Patla[35]	κ 0.40
Extension	Patla[35]	κ 0.73

κ: kappa

Elbow (n = 2)

Neither of the studies fulfilled all criteria for external or internal validity. Rothstein et al[36] demonstrated acceptable reliability for measuring range of flexion (ICC 0.85 to 0.97) and extension (0.92 to 0.95) using different types of goniometers in patients with elbow pathology. The reliability of measurements of physiological range of motion reported by Rothstein et al[36] was substantially higher than the reliability of judgements of end-feel of flexion (kappa 0.40) and extension (0.73) reported by Patla and Paris.[35]

Wrist-hand-fingers (n = 6)

One study[41] satisfied all criteria for internal validity. Almost perfect reliability (ICC 0.99, 95% CI 0.98 to 1.0), associated with a low risk of bias, was reported for measurements of passive torque-controlled physiological range of finger and thumb flexion/extension using a goniometer in patients with a traumatic hand injury.[41] Three studies[37-39] investigated the reliability of measurements of physiological range of motion at the wrist of which the latter two reported acceptable ICC values for wrist extension (0.80 to 0.84) and flexion (0.86 to 0.93) using goniometers. In contrast, Bovens et al[37] reported poor reliability for physicians using vision to measure physiological wrist extension. Reliability for measuring physiological thumb abduction was reported to be higher using a pollexograph (ICC 0.59, 95% CI 0.42 to 0.89) than a goniometer (0.37, 95% CI -0.42 to 0.79).[42] Finally, measuring accessory movements of carpal bones against the capitate bone using a 3-point scale yielded fair to moderate reliability (weighted kappa 0.29 to 0.42) in healthy individuals and fair to almost perfect (0.33 to 0.87) in post-operative patients.[40]

Discussion

This systematic review included 21 studies investigating inter-examiner reliability of measurements of passive movements of upper extremity joints of which 11 demonstrated acceptable reliability (ICC > 0.75). Reliability varied considerably with the method of measurement and ICC ranged from 0.26 (95% CI -0.01 to 0.69), for measuring the physiological range of internal shoulder rotation using vision, to 0.99 (95% CI 0.98 to 1.0), for the physiological range of finger and thumb flexion/extension using a goniometer. In general, measurements of physiological range of motion using instruments were more reliable than measurements using vision. Furthermore, measurements of physiological range of motion were also more reliable than judgements of end-feel or of accessory range of motion. Overall, methodological quality of included studies was poor, although two high-quality studies[30;41] reported almost perfect reliability.

In general, reliability for measurements of passive movements of upper extremity joints was substantially higher than that for judgements of passive segmental intervertebral or sacroiliac motion which rarely exceeds kappa 0.40.[15;16] Seffinger et al[13] attributed these differences in reliability to differences in size of joints. We believe, however, that differences may be more linked to a joint's potential physiological range of motion. For instance, measurements of large joints with limited range such as the sacroiliac joint is associated with poor reliability, whereas measurement of small joints with greater range such as the atlantoaxial spinal segment and the finger joints has been shown to be reliable.[16;41;43;44] We also found that measuring large physiological ranges of motion like those in the shoulder and in the wrist frequently yielded satisfactory levels of reliability and we note that these levels were predominantly a result of using goniometers or inclinometers. In addition, findings from four studies[23;26;34;35] indicated that judgements of end-feel or accessory movements of joints with large ranges of motion was associated with lower reliability. Staes et al[40], on the other hand, reported better reliability for end-feel assessment of accessory intercarpal motion as compared to mobility classifications. With respect to spinal movement, Haneline et al[10] similarly found somewhat higher reliability for judgements of end-feel. We hypothesise that measuring physiological movement for joints with large ranges of motion using goniometers or inclinometers, and judging end-feel for joints with limited range of motion will lead to more reliable decisions about joint restrictions in clinical practice. Since few studies have investigated reliability of judgements of end-feel or accessory movements in upper extremity joints, future research should focus on the inter-examiner reliability of these measures compared with measurements of physiological movements within the same sample of participants and examiners.

In this review, we found studies investigating inter-examiner reliability of upper extremity joint motion examination to have been poorly conducted. Only one study satisfied all external validity criteria and only two met all internal validity criteria. None of the included studies was both externally and internally valid. This finding is no different from that of reviews of the reliability of judgements of spinal motion.[13;15] The majority of the studies in our review met the criterion concerning blinding procedures. However, criteria about the stability of participants' and examiners' characteristics during the study were often either unmet or unknown. Instability of the participants' characteristics under investigation, in this case joint range of motion or end-feel, may be caused by changes in the biomechanical properties of connective tissues as a result of natural variation over time or the effect of the measurement procedure itself.[45] Similarly, instability of the examiners, in this case their consistency in making judgements, may be caused by mental fatigue. Instability of participants' or examiners' characteristics can lead to underestimations of reliability, whereas a lack of appropriate blinding of examiners can lead to overestimation. In the presence of all of these methodological flaws, the direction of risk of bias is difficult to predict. Factors about internal validity are closely linked to issues of generalisation of results. For instance, performing several measurements on a large number of participants in a limited time period is not only susceptible to bias but also does not reflect clinical practice.

The reliability of measurements varies across populations of participants and examiners.[7] In order to better reflect clinical practice, it is preferable to measure participants who would normally have their passive movements measured as part of the physiotherapy assessment, i.e., consecutive patients with musculoskeletal conditions rather than healthy volunteers, as well as allowing examiners access to information from history and physical examination as is usually gathered previous to passive motion examination.[19] However, we had decided *a priori* to include studies of asymptomatic individuals because of the information on reliability they may provide. Seven of our included studies used healthy volunteers as participants.

We note that the majority of included studies calculated ICC for expressing reliability of measurement of range of motion between examiners. ICC is the most appropriate parameter of reliability for continuous data reflecting the ability of examiners to discriminate between individuals.[8] For determining effects of intervention, however, insight into absolute measurement error is required and other parameters such as the limits of agreement are preferable for expressing agreement within examiners on measurements across multiple occasions over time.[8;46] To date, such data with respect to measurements of passive movements of upper extremity joints are rarely available. Since reliable measures of passive movement do not necessarily also have low levels of absolute measurement errors, they cannot necessarily be used when evaluating effects of intervention.

Finally, with regard to physiological range of motion in the shoulder, we found large variation in reliability of measurement of external rotation and abduction range. Cyriax[5] first described capsular patterns of joint restrictions to distinguish between capsular and other causes, e.g., external rotation being most limited followed by abduction followed by internal rotation. This pattern, however, was not corroborated in patients with idiopathic loss of shoulder range of motion.[47] In addition, almost complete loss of external rotation is the pathognomic sign of frozen shoulder.[48] Valid diagnosis of shoulder disorders based on pattern of passive external rotation and abduction loss of range requires further research.

Limitations of this study

This review has limitations with respect to its search strategy, quality assessment, and analysis. Only 11 included studies originated from our electronic search. A reason for this low electronic yield may be the inconsistent terminology used in reliability research. In our experience, reliability studies were poorly indexed in databases. In addition, our search strategy may have been too specific. Although much effort was put into reference tracing and hand searching, it is possible that eligible studies were missed. Furthermore, unpublished studies were not included. Publication bias can form a real threat to internal validity of systematic reviews of reliability studies, because they are more likely to report low reliability.

Quality assessment was performed by using criteria derived mainly from the quality assessment of diagnostic accuracy studies. No evidence is available on whether these criteria can be applied to reliability studies. Empirical evidence of bias, especially concerning blinding of examiners and stability of characteristics of participants and examiners, is lacking. Another method for scoring methodological quality may have resulted in different conclusions.

Finally, our analysis was based on point estimates of reliability. Including interpretation of the precision of these estimates would have provided a more detailed perspective. However, only a limited number of included studies presented 95% CI. In the majority of these cases, CI were quite wide suggesting low sample sizes. None of our included studies reported an *a priori* sample size calculation.

Conclusions and recommendations

We conclude that the inter-examiner reliability of measurements of passive movements in upper extremity joints varies with the method of assessment. In order to make reliable decisions about joint restriction in clinical practice, we recommend that clinicians measure passive physiological range of motion using goniometers or inclinometers. Future research should focus on comparing inter-examiner reliability of end-feel and accessory movements with passive physiological range of motion assessment, using symptomatic individuals. In addition, more research is needed on the elbow and wrist joints. Careful consideration should be given to ensuring stability of participants' and examiners' characteristics during the study and *a priori* sample sizes should be calculated. Following the STARD statement will also improve the quality of reporting of reliability studies.[17;18] Finally, new intra-examiner reliability studies determining the absolute measurement error (agreement) when measuring passive range of motion in upper extremity joints will provide insight into the amount of change in range needed to an effect of intervention beyond this error.

References

1. Ho C-YC, Sole G, Munn J. The effectiveness of manual therapy in the management of musculoskeletal disorders of the shoulder: A systematic review. Man Ther 2009;14:463-74.
2. Kromer TO, Tautenhahn UG, De Bie RA, Staal JB, Bastiaenen CHG. Effects of physiotherapy in patients with shoulder impingement syndrome: a systematic review of the literature. J Rehabil Med 2009;41:870-80.
3. Bisset L, Paungmali A, Vicenzino B, Beller E. A systematic review and meta-analysis of clinical trials on physical interventions for lateral epicondylalgia. Br J Sports Med 2005; 39:411–22.
4. O'Connor D, Marshall SC, Massy-Westropp N. Non-surgical treatment (other than steroid injection) for carpal tunnel syndrome. Cochrane Database of Systematic Reviews 2003, Issue 1. Art. No.: CD003219.
5. Cyriax J. Textbook of orthopaedic medicine. Volume one: Diagnosis of soft tissue lesions (8^{th} edition). London: Bailliere Tindall, 1982.
6. Hengeveld E, Banks K. Maitland's peripheral manipulation (4^{th} edition). Philadelphia: Butterworth-Heinemann, 2005.
7. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (4^{th} edition). Oxford, UK: Oxford University Press, 2008.
8. De Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
9. Bartko JJ, Carpenter WT (1976) On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
10. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
11. Hestbæk L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther 2000;23:258-75.
12. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Aust J Physiother 2006;52:91-102.

13. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Spine 2004;29:E413-25.
14. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbaek L, Adams A, Bronfort G. Manual examination of the spine: a systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85.
15. Van Trijffel E, Anderegg Q, Bossuyt PM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. Man Ther 2005;10:256-69.
16. Van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. Man Ther 2000;5:30-6.
17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. Clin Chem 2003;49:1-6.
18. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, De Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
19. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
20. Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111-23.
21. Landis JR, Koch DG. The measurement of observer agreement for categorical data. Biometrica 1977;33:159–164.
22. Awan R, Smith J, Boon AJ. Measuring shoulder internal rotation range of motion: a comparison of 3 techniques. Arch Phys Med Rehabil 2002;83:1229-34.
23. Chesworth BM, MacDermid JC, Roth JH, Patterson SD. Movement diagram and "end-feel" reliability when measuring passive lateral rotation of the shoulder in patients with shoulder pathology. Phys Ther 1998;78:593-601.
24. De Winter AF, Heemskerk MA, Terwee CB, Jans MP, Devillé W, Van Schaardenburg DJ, Scholten RJ, Bouter LM. Inter-observer reproducibility of measurements of range of

motion in patients with shoulder pain using a digital inclinometer. BMC Musculoskelet Disord 2003;5:18.

25. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47:289-94.
26. Hayes KW, Petersen CM. Reliability of assessing end-feel and pain and resistance sequence in subjects with painful shoulders and knees. J Orthop Sports Phys Ther 2001;31:432-45.
27. Heemskerk MAMB, Van Aarst M, Van der Windt DAWM. De reproduceerbaarheid van het meten van de passieve beweeglijkheid van de schouder met de EDI-320 digitale hoekmeter. Dutch Journal of Physiotherapy 1997;107:146-9. [In Dutch]
28. Lin JJ, Yang JL. Reliability and validity of shoulder tightness measurement in patients with stiff shoulders. Man Ther 2006;11:146-52.
29. MacDermid JC, Chesworth BM, Patterson S, Roth JH. Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. J Hand Ther 1999;12:187-92.
30. Nomden JG, Slagers AJ, Bergman GJ, Winters JC, Kropmans TJ, Dijkstra PU. Interobserver reliability of physical examination of shoulder girdle. Man Ther 2009;14:152-9.
31. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting. Shoulder measurements. Phys Ther 1987;67:668-73.
32. Terwee CB, De Winter AF, Scholten RJ, Jans MP, Devillé W, Van Schaardenburg D, Bouter LM. Interobserver reproducibility of the visual estimation of range of motion of the shoulder. Arch Phys Med Rehabil 2005;86:1356-61.
33. Tyler TF, Roy T, Nicholas SJ, Gleim GW. Reliability and validity of a new method of measuring posterior shoulder tightness. J Orthop Sports Phys Ther 1999;29:262-9.
34. Van Duijn AJ, Jensen RH. Reliability of inferior glide mobility testing of the glenohumeral joint. J Man Manip Ther 2001;9:109-114.
35. Patla CE, Paris SV. Reliability of interpretation of the Paris classification of normal end feel for elbow flexion and extension. J Man Manip Ther 1993;1:60-66.
36. Rothstein JM, Miller PJ, Roettger RF. Goniometric reliability in a clinical setting. Elbow and knee measurements. Phys Ther 1983;63:1611-5.
37. Bovens AM, Van Baak MA, Vrencken JG, Wijnen JA, Verstappen FT. Variability and reliability of joint measurements. Am J Sports Med 1990;18:58-63.

38. Horger MM. The reliability of goniometric measurements of active and passive wrist motions. Am J Occup Ther 1990;44:342-8.
39. LaStayo PC, Wheeler DL. Reliability of passive wrist flexion and extension goniometric measurements: a multicenter study. Phys Ther 1994;74:162-74.
40. Staes FF, Banks KJ, De Smet L, Daniels KJ, Carels P. Reliability of accessory motion testing at the carpal joints. Man Ther 2009;14:292-8.
41. Glasgow C, Wilton J, Tooth L. Optimal daily total end range time for contracture: resolution in hand splinting. J Hand Ther 2003;16:207-18.
42. De Kraker M, Selles RW, Schreuders TAR, Stam HJ, Hovius SER. Palmar abduction: Reliability of 6 measurement methods in healthy adults. J Hand Surg Am 2009;34:523-30.
43. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 2006;87:1388-95.
44. Ogince M, Hall T, Robinson K, Blackmore AM. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headache. Man Ther 2007;12:256-62.
45. Rothstein JM, Echternach JL. Primer on measurement: an introductory guide to measurement issues. Alexandria, VA: American Physical Therapy Association, 1993.
46. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. Lancet 1986;1:307-10.
47. Rundquist PJ, Ludewig PM. Patterns of motion loss in subjects with idiopathic loss of shoulder range of motion. Clin Biomech 2004;19:810-8.
48. Dias R, Cutts S, Massoud S. Frozen shoulder. BMJ 2005;331:1453-6.

Appendix 1

Search strategy for MEDLINE (PubMed) using text words [tw] and Medical Subject Headings [mh]

1. shoulder [mh]
2. shoulder [tw]
3. humeral [tw]
4. glenohumeral [tw]
5. acromioclavicular [tw]
6. sternoclavicular [tw]
7. elbow [mh]
8. elbow [tw]
9. humeroulnar [tw]
10. humeroradial [tw]
11. radioulnar [tw]
12. wrist [mh]
13. wrist [tw]
14. carpal [tw]
15. radiocarpal [tw]
16. midcarpal [tw]
17. intercarpal [tw]
18. hand [mh]
19. hand [tw]
20. carpometacarpal [tw]
21. metacarpophlangeal [tw]
22. finger* [mh]
23. finger* [tw]
24. thumb [mh]
25. thumb [tw]
26. phalangeal [tw]
27. interphalangeal [tw]

28. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27
29. 28 AND Joint* [mh]
30. motion [mh]
31. movement [mh]
32. range of motion, articular [mh]
33. mobility [tw]
34. endfeel [tw]
35. end feel [tw]
36. 30 OR 31 OR 32 OR 33 OR 34 OR 35
37. physical examination [mh]
38. diagnostic tests, routine [mh]
39. observation [mh]
40. passive [tw]
41. manual [tw]
42. 37 OR 38 OR 39 OR 40 OR 41
43. reproducibility of results [mh]
44. reproducibility [tw]
45. reliability [tw]
46. observer variation [tw]
47. repeatability [tw]
48. variation [tw]
49. concordance [tw]
50. variability [tw]
51. agreement [tw]
52. 43 OR 44 OR 45 OR 46 OR 47 OR 48 OR 49 OR 50 OR 51
53. interobserver [tw]
54. intertester [tw]
55. interrater [tw]
56. interexaminer [tw]
57. observer* [tw]
58. tester* [tw]

59. rater* [tw]

60. examiner* [tw]

61. 53 OR 454 OR 55 OR 56 OR 57 OR 58 OR 59 OR 60

62. 29 AND 36 AND 42 AND 52 AND 61

Appendix 2

Excluded studies *(n = 25)* with their main reason for exclusion

Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L. Reliability of goniometric measurements. Phys Ther 1978;58:1355-90.
Reason for exclusion: Evaluating active range of motion only

Borstad JD, Mathiowetz KM, Minday LE, Prabhu B, Christopherson DE, Ludewig PM. Clinical measurement of posterior shoulder flexibility. Man Ther 2007;12:386-9.
Reason for exclusion: Evaluating intra-examiner reliability only

Boström C, Harms-Ringdahl K, Nordemar R. Clinical reliability of shoulder function assessment in patients with rheumatoid arthritis. Scand J Rheumatol 1991;20:36-48.
Reason for exclusion: Full text missing

Croft AC, Krage JS, Pate D, Young DN. Videofluoroscopy in cervical spine trauma: an interinterpreter reliability study. J Manipulative Physiol Ther 1994;17:20-4.
Reason for exclusion: Evaluating active range of motion only

De Jong LD, Nieuwboer A, Aufdemkampe G. The hemiplegic arm: interrater reliability and concurrent validity of passive range of motion measurements. Disabil Rehabil 2007;29:1442-8.
Reason for exclusion: Including patients with neurological deficit

Dijkstra PU, De Bont LG, Van der Weele LT, Boering G. Joint mobility measurements: reliability of a standardized method. Cranio 1994;12:52-7.
Reason for exclusion: Evaluating active range of motion only

Erkula G, Kiter AE, Kilic BA, Er E, Demirkan F, Sponseller PD. The relation of joint laxity and trunk rotation. J Pediatr Orthop B 2005;14:38-41.
Reason for exclusion: Evaluating active range of motion only

Flowers KR, Stephens-Chisar J, LaStayo P, Galante BL. Intrarater reliability of a new method and instrumentation for measuring passive supination and pronation: a preliminary study. J Hand Ther 2001;14:30-5.
Reason for exclusion: Evaluating intra-examiner reliability only

Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther 1987;67:1867-72.
Reason for exclusion: Review study

Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. BMJ 1998;316:354-360.
Reason for exclusion: Review study

Greene BL, Wolf SL. Upper extremity joint movement: comparison of two measurement devices. Arch Phys Med Rehabil 1989;70:288-90.
Reason for exclusion: Evaluating active range of motion only

Gross ML, Distefano MC. Anterior release test. A new test for occult shoulder instability. Clin Orthop Relat Res 1997;339:105-8.
Reason for exclusion: Evaluating joint stability

Levy AS, Lintner S, Kenter K, Speer KP. Intra- and interobserver reproducibility of the shoulder laxity examination. Am J Sports Med 1999;27:460-3.
Reason for exclusion: Evaluating joint stability

Lin HT, Hsu AT, An KN, Chang Chien JR, Kuan TS, Chang GL. Reliability of stiffness measured in glenohumeral joint and its application to assess the effect of end-range mobilization in subjects with adhesive capsulitis. Man Ther 2008;13:307-16.
Reason for exclusion: Using instruments not feasible in practice

Lo IK, Nonweiler B, Woolfrey M, Litchfield R, Kirkley A. An evaluation of the apprehension, relocation, and surprise tests for anterior shoulder instability. Am J Sports Med 2004;32:301-7.
Reason for exclusion: Evaluating joint stability

Loessin Grohmann JE. Comparison of two methods of goniometry. Phys Ther 1983;63:922-5.
Reason for exclusion: Full text missing

Low JL. The reliability of joint measurement. Physiotherapy 1976;62:227-9.
Reason for exclusion: Full text missing

Mayerson NH, Milano RA. Goniometric measurement reliability in physical medicine. Arch Phys Med Rehabil 1984;65:92-4.
Reason for exclusion: Full text missing

McLauchlan GJ, Walker CR, Cowan B, Robb JE, Prescott RJ. Extension of the elbow and supracondylar fractures in children. J Bone Joint Surg Br 1999;81:402-5.
Reason for exclusion: Evaluating active range of motion only

Piotte F, Gravel D, Nadeau S, Moffet H, Bédard C. Reliability of arthrometric measurement of shoulder lateral rotation movement in healthy subjects. Physiother Theory Pract 2007;23:169-78.
Reason for exclusion: Using instruments not feasible in practice

Pohl M, Mehrholz J. A new shoulder range of motion screening measurement: its reliability and application in the assessment of the prevalence of shoulder contractures in patients with impaired consciousness caused by severe brain damage. Arch Phys Med Rehabil 2005;86:98-104.
Reason for exclusion: Including patients with neurological deficit

Solgaard S, Carlsen A, Kramhøft M, Petersen VS. Reproducibility of goniometry of the wrist. Scand J Rehabil Med 1986;18:5-7.
Reason for exclusion: Evaluating active range of motion only

Tillander B, Norlin R. Intraoperative measurement of shoulder translation. J Shoulder Elbow Surg 2001;10:358-64.
Reason for exclusion: Evaluating joint stability

Van de Ende CH, Rozing PM, Dijkmans BA, Verhoef JA, Voogt-van der Harst EM, Hazes JM. Assessment of shoulder function in rheumatoid arthritis. J Rheumatol 1996;23:2043-8.
Reason for exclusion: Evaluating active range of motion only

Williams JG, Callaghan M. Comparison of visual estimation and goniometry in determination of a shoulder joint angle. Physiotherapy 1990;76:655-7.
Reason for exclusion: Full text missing

Chapter 3b

Inter-examiner reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review

Emiel van Trijffel, Rachel van de Pol, Rob Oostendorp, Cees Lucas

J Physiother 2010;56:223-35

Abstract

Background: Passive assessment of motion in joints of the lower extremity is commonly used by physiotherapists in order to measure joint restrictions and to diagnose musculoskeletal disorders. To date, no systematic appraisal of studies on the inter-examiner reliability of measurement of passive movements in lower extremity joints has been conducted.

Methods: We conducted a systematic review of studies of inter-examiner reliability of measurements of passive movement in lower extremity joints published up to March 1, 2010. Studies involving participants with and without lower extremity disorders were included. Range of motion measurements and end-feel judgements from passive joint motion examination using methods and instruments feasible in clinical practice were considered. No language restrictions were imposed. Study selection, quality assessment, and data extraction were performed by two reviewers independently.

Results: Seventeen studies were included of which five demonstrated acceptable inter-examiner reliability. Reliability of measurements of physiological range of motion ranged from kappa -0.02, for measuring knee extension using a goniometer to ICC 0.97, for measuring knee flexion using vision. Measuring range of knee flexion consistently yielded acceptable reliability using either vision or instruments. Judgements of end-feel were unreliable for all hip and knee movements. Two studies satisfied all criteria for internal validity while reporting acceptable reliability for measuring physiological range of knee flexion and extension. Overall, however, methodological quality of included studies was poor.

Conclusions: Inter-examiner reliability of measurement of passive movements in lower extremity joints is generally low. We provide specific recommendations for the conduct and reporting of future research. Awaiting new evidence, clinicians should be cautious when relying on results from measurements of passive movements in joints for making decisions about patients with lower extremity disorders.

Introduction

Physiotherapists commonly assess and treat patients with lower extremity joint disorders. Despite varying levels of evidence, a growing number of studies have shown that manual joint mobilisation or manipulation is effective in certain disorders such as hip and knee osteoarthritis, patellofemoral pain syndrome, ankle inversion sprain, plantar fasciitis, metatarsalgia, and hallux limitus/rigidus.[1] Measurement of passive movement is indicated in order to assess joint restrictions and to help diagnose these disorders. Passive movement, either physiological or accessory, can be reported as range of motion, end-feel, or pain and is an indication of the integrity of joint structures.[2-4] Passive physiological range of motion may be measured using vision or instruments such as goniometers or inclinometers.

An essential requirement of clinical measures is that they are valid and reliable so that they can be used to discriminate between individuals.[5] Inter-examiner reliability is a component of reproducibility along with agreement and refers to the relative measurement error, i.e., the variation between patients as measured by different examiners in relation to the total variance of the measurements.[5;6] High inter-examiner reliability for measurements of lower extremity joints is a prerequisite for valid and uniform clinical decisions about joint restrictions and related disorders.[7]

Several reviews have systematically summarised and appraised the evidence with respect to the inter-examiner reliability of passive movements of human joints. Seven systematic reviews have been published on passive spinal and pelvic movement including segmental intervertebral motion assessment.[8-14] In general, inter-examiner reliability was found to be poor and studies were of low methodological quality. A recent systematic review showed better inter-examiner reliability for measurements of passive physiological range of motion in upper extremity joints using instruments compared to measurements using vision and compared to measurements of end-feel or accessory range of motion.[15] To date, no systematic appraisal of studies on inter-examiner reliability of measurement of passive movements in lower extremity joints has been conducted. Therefore, the research question for this systematic review was: What is the inter-examiner reliability for measurements of passive physiological or accessory movements in lower extremity joints?

Methods

Study selection

MEDLINE, EMBASE, and CINAHL were searched for studies published up to March 1, 2010. Search terms included all lower extremity joints and all synonyms for *reliability* and *examiner* (Appendix 1). The titles and abstracts were screened for eligibility by two reviewers (EvT, RvdP) independently. When necessary, full text articles were retrieved. Reference lists of all retrieved papers were hand searched for relevant studies. A supplemental hand search of 13 journals relevant to the field of physiotherapy from January 1, 2005, to March 1, 2010, was performed by one reviewer (EvT). A complete list of journals is available from the authors. Finally, four experts in lower extremity musculoskeletal research were approached to ask if they could provide any additional published studies. Additionally retrieved papers were checked for eligibility by a second reviewer (RvdP). Studies were included if they met all inclusion criteria (Box 1).

Box 1. Inclusion criteria

Design

- Repeated measures between examiners

Participants

- Symptomatic and asymptomatic individuals

Measurement procedure

- Performed passive (i.e., manual) physiological or accessory movements in any of the joints of the hip, knee, or ankle-foot-toes
- Reported range of motion or end-feel
- Used methods feasible in clinical practice (considering instruments, costs, amount of training required)

Outcomes

- Estimates of inter-examiner reliability

No restrictions were imposed on language or date of publication. Studies were excluded if they were abstracts or documents that were anecdotal, speculative, or editorial in nature. Studies were also excluded if they investigated: active movement or restriction in passive movement due to pain or ligament instability; people with neurological conditions in which abnormal muscle tone may interfere with joint movement; people after arthroplasty; animals or cadavers. Study selection was performed by two reviewers (EvT, RvdP) independently. Disagreements on eligibility were first resolved by discussion between the two reviewers and decided by a third reviewer (CL) if disagreement persisted.

Quality assessment

No validated instrument was available for assessing methodological quality of inter-examiner reliability studies. Therefore, a list of criteria for quality was compiled derived from the QUADAS tool, the STARD statement, and criteria used for assessing studies on reliability of measuring passive spinal movement.[14;16-18] Criteria 1 to 4 assess external validity, Criteria 5 to 9 assess internal validity, and Criterion 10 assesses statistical methods (Box 2).

Box 2. Criteria for assessing methodological quality

1. Was a representative sample of participants used?
2. Was a representative sample of examiners used?
3. Is replication of the measurement procedure possible?
4. Was clinical information from participants available to examiners and comparable to clinical practice?
5. Were participants' characteristics stable during the study?
6. Were examiners' characteristics stable during the study?
7. Were examiners blinded to each other's results?
8. Can non-random loss to follow-up be ruled out?
9. Was an estimate of intra-examiner reliability validly determined and was it above 0.80?
10. Were appropriate measures (kappa, intraclass correlation coefficient) used for calculating reliability?

Criteria were rated as 'Yes', 'No', or '?' (unclear because of insufficient information). External validity was considered sufficient if Criteria 1 to 4 were rated 'Yes'. With respect to internal validity, Criteria 5, 6, and 7 were assumed to be decisive in determining risk of bias. A study was considered to have a low risk of bias if Criteria 5, 6, and 7 were all rated 'Yes', a moderate risk if two of these criteria were rated 'Yes', and a high risk if none or only one of these criteria were rated 'Yes'. After training, two reviewers (EvT, RvdP) independently assessed methodological quality of all included studies and were not blind to journal, authors, and results. If discrepancy between reviewers persisted, a decisive judgement was passed by a third reviewer (CL).

Data extraction

We extracted data on participants (number, age, clinical characteristics), examiners (number, profession, training), measurements (joints and movement direction, participant position, movement performed, method of measurement, outcomes reported), and inter-examiner reliability (point estimates, estimates of precision). Two reviewers (EvT, RvdP) extracted data independently and were not blind to journal, authors, or results. When disagreement between the two reviewers could not be resolved by discussion, a third reviewer (CL) made the final decision.

Data analysis

Data were analysed by examining intraclass correlation coefficient (ICC) and Cohen's kappa (95% CI). If at least 75% of a study's ICC or kappa values were above 0.75, the study was considered to have shown acceptable reliability.[19] Corresponding kappa levels were used as assigned by Landis & Koch where <0.00 = poor, 0.00-0.20 = slight, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = substantial, and 0.81-1.00 = almost perfect reliability.[20] In addition, reliability was analysed relating it to characteristics of the studies (participants' clinical characteristics, examiners' profession and training, movement performed, method of measurement) and methodological quality. Reliability from studies not fulfilling Criteria 5 or 6 could have been underestimated while reliability from studies not fulfilling Criterion 7 could have been overestimated. Negative scores on combinations of Criteria 5-7 could have led to bias in an unknown direction. Where one or more of these three criteria were unclear, no statement was made regarding the presence or direction of potential bias. Finally, clinical and methodological characteristics of included studies were examined for homogeneity in order to judge the possibility of statistically summarising results by calculating pooled estimates of reliability.

Results

Flow of studies through the review

Searching MEDLINE yielded 199 citations of which 29 papers were retrieved in full text. After removing double citations, EMBASE (196 citations) provided another three potentially relevant studies. CINAHL (98 citations) then yielded no additional relevant articles. Hand searching of reference lists identified another 14 potentially eligible studies. Of these 46, 31 studies were excluded (Appendix 2). Hand searching of journals yielded one eligible study while one expert provided another. In total, 17 studies fulfilled all inclusion criteria (Figure).

Characteristics of included studies

The included studies are summarised in Table 1. Seven studies[21-27] investigated inter-examiner reliability of measurement of passive hip movements, seven[25;28-33] investigated knee movements, five[27;34-37] investigated ankle movements, and one[27] investigated first ray movements. In 11 studies, physiotherapists acted as examiners. There were no disagreements between reviewers on selection of studies.

Quality of studies

The methodological quality of included studies is presented in Table 2.

One study[37] fulfilled all four criteria for external validity and four studies[23;31;33;35] satisfied three criteria. Two studies[28;33] fulfilled all three criteria for internal validity representing a low risk of bias while five studies[23;30;34;35;37] satisfied two criteria. Criteria on external and internal validity could not be scored on 49/153 (32%) occasions because of insufficient reporting. On methodological quality scores, 12/170 (7%) disagreements occurred between reviewers which were all resolved by discussion.

Inter-examiner reliability by region

The inter-examiner reliability for measurements of physiological range of motion is presented in Table 3, and for judgements of physiological end-feel in Table 4. Because of clinical and methodological heterogeneity between studies, we did not attempt to calculate pooled estimates of reliability.

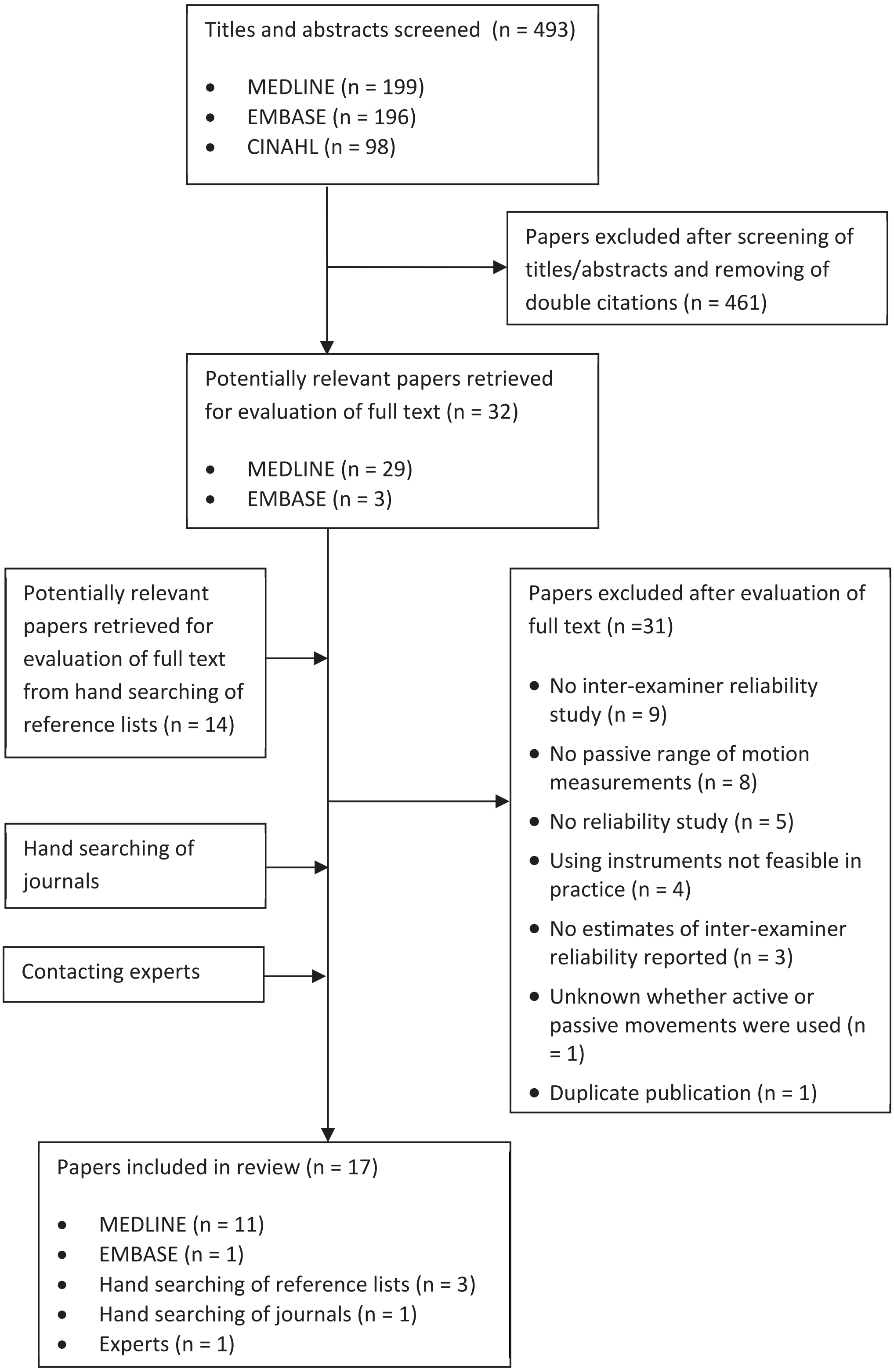

Figure. Flow of studies through the review

Table 1. Characteristics of included studies *(n = 17)*

Study	Participants	Examiners	Joints Movement directions	Position	Movement performed	Method	Outcome reported	Reliability statistic
Aalto[21]	n = 20 Age = mean 23.3 yr (range 18-45) Condition = normal	n = 2 Profession = PT Training = ?	Hip • IR	 Seated	Physiological	Goniometer	ROM	ICC
Chevillotte[22]	n = 33 Age = mean 62.8 yr (SD 16.1) Condition = preoperative hip OA	n = 5 Profession = 2 hip surgeons, 2 orthopaedic surgery residents, 1 physician assistant Training = N	Hip • F • Abd • Add • IR • ER • E	 Supine Supine Hip 90 deg F Lateral decubitus	Physiological	Vision	ROM	ICC
Cibere[28]	n = 6 Age = median 62 yr (range 44-74) Condition = knee OA	n = 6 Profession = rheumatologist Training = Y	Knee • E	 ?	Physiological	Goniometer	ROM	PABAK
Cibere[23]	n = 6 Age = median 63 yr (range 49-65) Condition = hip OA	n = 6 Profession = 4 rheumatologists, 2 orthopaedic surgeons Training = Y	Hip • F • Abd • Add • IR • ER • IR • ER • E	 Supine Supine Hip 90 deg F Seated Lateral decubitus	Physiological	Goniometer	ROM	R

Cleffken[29]	n = 42 Age = mean 22.1 yr (range 19-27) Condition = normal	n = 2 Profession = ? Training = Y	Knee • F	Supine	Physiological	Inclinometer	ROM	Pearson's r
Croft[24]	n = 6 Age = ? Condition = hip OA	n = 6 Profession = 5 general practitioner, 1 hospital physician Training = Y	Hip • F • IR • ER	Supine Seated	Physiological	Plurimeter	ROM	ICC
Currier[25]	n = 25 Age = ? Condition = knee OA	n = 2 Profession = PT doctoral student Training = Y	Hip • F • E • Abd • Add • Distraction • Patrick's test • IR • ER Knee • F • E	Supine Prone Knee 90 deg F Supine	Physiological	Goniometer Inclinometer	ROM End-feel	ICC (2,1) kappa
Diamond[34]	n = 31 Age = mean 59 yr (SD 12) Condition = diabetes mellitus	n = 2 Profession = PT Training = Y	Ankle • DF • INV • EV	Prone Knee 0 deg F	Physiological	Goniometer	ROM	ICC (2,1)
Elveru[35]	n = 43 Age = mean 35.9 yr (SD 15.6) Condition = general orthopaedic disorders	n =14 Profession = PT Training = Y	Ankle • DF • PLF • INV • EV	Prone Knee 0 deg F	Physiological	Goniometer	ROM	ICC (1,1)

Erichsen[36]	n = 27 Age = range 20-45 yr Condition = ankle pathology, normal	n = 2 Profession = PT Training = Y	Ankle • PLF • INV-EV • Med-lat talus glide	Supine	Physiological Accessory	Vision	ROM	kappa
Fritz[30]	n = 35 Age = ? Condition = knee dysfunction	n = 9 Profession = PT Training = N	Knee • F	Supine	Physiological	Vision	ROM	ICC (2,1)
Hayes[31]	n = 17 Age = mean 31.8 yr (SD 9.5) Condition = knee pain	n = 2 Profession = PT Training = Y	Knee • F • E	Supine	Physiological	Manual	End-feel	kappa
Rothstein[32]	n = 12 Age = ? Condition = knee pathology	n = 12 Profession = PT Training = ?	Knee • F • E	?	Physiological	Goniometer	ROM	ICC
Smith-Oricchio[37]	n = 20 Age = range 18-53 yr Condition = ankle pathology	n = 3 Profession = PT Training = N	Ankle • INV • EV	Prone Knee 0 deg F	Physiological	Goniometer	ROM	ICC (3,1)
Sutlive[26]	n = 30 Age = ? Condition = hip pain	n = ? Profession = PT doctoral student Training = Y	Hip • IR • ER • Flexion • Scour test • Patrick's test	Prone Knee 90 deg F Supine	Physiological	Inclinometer	ROM End-feel	ICC (2,1) kappa

Van Gheluwe[27]	n = 30 Age = mean 24.8 yr Condition = normal	n = 5 Profession = podiatric physician Training = N	Hip • IR • ER Ankle • DF • INV • EV First ray • DF • PLF	?	Physiological	Goniometer	ROM	Two-way ICC
Watkins[33]	n = 43 Age = mean 39.5 yr (SD 15) Condition = knee pathology	n = 14 Profession = PT Training = N	Knee • F • E	?	Physiological	Vision Goniometer	ROM	ICC (1,1)

Abd: abduction, Add: adduction, DF: dorsiflexion, EV: eversion, ER: external rotation, E: extension, F: flexion, ICC: intraclass correlation coefficient, IR: internal rotation, INV: inversion, OA: osteoarthritis, PLF: plantar flexion, PABAK: prevalence-adjusted bias-adjusted kappa, PT: physiotherapist, R: correlation coefficient, ROM: range of motion, THA: total hip arthroplasty

Table 2. Quality of studies *(n = 17)*

	External validity				Internal validity					Statistical methods
Study	1	2	3	4	5	6	7	8	9	10
Aalto[21]	N	?	Y	N	N	?	?	Y	Y	?
Chevillotte[22]	Y	?	N	?	N	N	?	Y	N	?
Cibere[28]	Y	?	N	N	Y	Y	Y	Y	?	Y
Cibere[23]	Y	Y	Y	?	?	Y	Y	Y	?	Y
Cleffken[29]	N	?	Y	?	Y	?	?	Y	N	N
Croft[24]	?	?	Y	?	?	Y	?	Y	?	?
Currier[25]	N	N	N	N	?	?	Y	Y	?	Y
Diamond[34]	Y	?	Y	?	Y	N	Y	Y	N	Y
Elveru[35]	Y	Y	Y	?	Y	Y	?	Y	N	Y
Erichsen[36]	Y	N	Y	?	?	N	Y	Y	N	Y
Fritz[30]	Y	?	Y	?	Y	N	Y	Y	?	Y
Hayes[31]	N	Y	Y	Y	?	?	Y	Y	N	?
Rothstein[32]	?	Y	?	N	?	?	Y	Y	N	?
Smith-Oricchio[37]	Y	Y	Y	Y	Y	?	Y	Y	?	Y
Sutlive[26]	N	N	Y	?	?	?	Y	Y	?	Y
Van Gheluwe[27]	N	Y	Y	?	?	?	?	Y	N	Y
Watkins[33]	Y	Y	N	Y	Y	Y	Y	Y	N	Y

Y: Yes, N: No, ?: unclear because of insufficient information

Hip (n = 7)

None of the studies fulfilled all criteria for external or internal validity. In two studies[21;23], acceptable reliability was reached. Inter-examiner reliability (ICC) of measurements of passive physiological range of motion ranged from 0.12 (95% CI 0.00 to 0.35), for surgeons and a physician assistant using vision to measure extension in preoperative patients with hip osteoarthritis[22], to 0.91, for physiotherapists using a goniometer to measure internal rotation in non-symptomatic participants[21]. Chevillotte et al[22] found unacceptable reliability for measurements of all physiological hip movements. However, their estimates could have been underestimated due to instability of characteristics of participants as well as of

Table 3. Inter-examiner reliability (95% CI) of passive physiological range of motion by method of measurement, joint, and movement direction

Method of measurement	Study	Inter-examiner reliability
Goniometer		
Hip		
Flexion	Cibere[23]	R = 0.91, 0.91
Extension	Cibere[23]	R = 0.66
Internal rotation	Aalto[21]	ICC = 0.75 to 0.91
	Cibere[23]	R = 0.87 to 0.95
	Van Gheluwe[27]	ICC = 0.41 (lower limit 0.26) to 0.51 (lower limit 0.35)
External rotation	Cibere[23]	R = 0.55 to 0.87
	Van Gheluwe[27]	ICC = 0.35 (lower limit 0.20) to 0.37 (lower limit 0.21)
Abduction	Cibere[23]	R = 0.88, 0.91
	Currier[25]	ICC = 0.54 (0.19 to 0.76)
Adduction	Cibere[23]	R = 0.56, 0.72
	Currier[25]	ICC = 0.37 (-0.03 to 0.67)
Knee		
Flexion	Currier[25]	ICC = 0.87 (0.73 to 0.94)
	Rothstein[32]	ICC = 0.84 to 0.93
	Watkins[33]	ICC = 0.90
Extension	Cibere[28]	PABAK = -0.02, 0.88
	Currier[25]	ICC = 0.69 (0.41 to 0.85)
	Rothstein[32]	ICC = 0.59 to 0.80
	Watkins[33]	ICC = 0.86
Ankle		
Dorsiflexion	Diamond[34]	ICC = 0.74, 0.87
	Elveru[35]	ICC = 0.00
	Van Gheluwe[27]	ICC = 0.26 (lower limit 0.12), 0.31 (lower limit 0.17)
Plantar flexion	Elveru[35]	ICC = 0.74
Inversion	Diamond[34]	ICC = 0.86, 0.88
	Elveru[35]	ICC = 0.30
	Smith-Oricchio[37]	ICC = 0.42
	Van Gheluwe[27]	ICC = 0.28 (lower limit 0.14), 0.40 (lower limit 0.22)
Eversion	Diamond[34]	ICC = 0.78, 0.79
	Elveru[35]	ICC = 0.22

	Smith-Oricchio[37]	ICC = 0.25
	Van Gheluwe[27]	ICC = 0.46 (lower limit 0.30), 0.49 (lower limit 0.32)
First ray		
Dorsiflexion	Van Gheluwe[27]	ICC = 0.14 (lower limit 0.04), 0.16 (lower limit 0.06)
Plantar flexion	Van Gheluwe[27]	ICC = 0.19 (lower limit 0.07), 0.21 (lower limit 0.09)
Vision		
Hip		
Flexion	Chevillotte[22]	ICC = 0.56 (0.37 to 0.75)
Extension	Chevillotte[22]	ICC = 0.12 (0.00 to 0.35)
Internal rotation	Chevillotte[22]	ICC = 0.50 (0.30 to 0.70)
External rotation	Chevillotte[22]	ICC = 0.37 (0.19 to 0.60)
Abduction	Chevillotte[22]	ICC = 0.49 (0.29 to 0.70)
Adduction	Chevillotte[22]	ICC = 0.39 (0.20 to 0.62)
Knee		
Flexion	Fritz[30]	ICC = 0.97
	Watkins[33]	ICC = 0.83
Extension	Watkins[33]	ICC = 0.82
Ankle		
Plantar flexion	Erichsen[36]	κ = 0.20 (-0.22 to 0.63), 0.47 (0.13 to 0.81)
Inversion-eversion	Erichsen[36]	κ = 0.37 (-0.03 to 0.77), 0.37 (-0.03 to 0.77)
Inclinometer		
Hip		
Flexion	Currier[25]	ICC = 0.56 (0.21 to 0.78)
Extension	Currier[25]	ICC = 0.20 (-0.22 to 0.55)
Internal rotation	Currier[25]	ICC = 0.76 (0.53 to 0.89)
	Sutlive[26]	ICC = 0.88 (0.74 to 0.94)
External rotation	Currier[25]	ICC = 0.29 (-0.12 to 0.62)
	Sutlive[26]	ICC = 0.77 (0.53 to 0.89)
Patrick's test	Currier[25]	ICC = 0.57 (0.23 to 0.79)
Knee		
Flexion	Cleffken[29]	Pearson's r = 0.83 to 0.87

Plurimeter		
Hip		
Flexion	Croft[24]	ICC = 0.87
Internal rotation	Croft[24]	ICC = 0.48
External rotation	Croft[24]	ICC = 0.43

ICC: intraclass correlation coefficient, κ: kappa, PABAK: prevalence-adjusted bias-adjusted kappa, R: correlation coefficient

examiners. Cibere et al[23] found acceptable reliability for measuring range of flexion, abduction, and internal rotation using a goniometer by trained rheumatologists and orthopaedic surgeons in patients with hip osteoarthritis. No specific movement direction or method of measurement was consistently associated with high or low reliability. Inter-examiner reliability (kappa) of judgements of physiological end-feel ranged from poor (-0.13, 95% CI -0.48 to 0.22) for extension[25], to moderate (0.52, 95% CI 0.08 to 0.96), for the Scour test[26]. Both studies[25;26] investigating reliability of end-feel judgements used symptomatic participants.

Knee (n = 7)

Two studies[28;33] fulfilled all criteria for internal validity. Cibere et al[28] demonstrated almost perfect inter-examiner reliability (kappa 0.88) for rheumatologists using a goniometer to measure passive physiological range of extension in patients with knee osteoarthritis. Watkins et al[33] reported acceptable reliability for physiotherapists using either vision or a goniometer to measure physiological range of flexion and extension in symptomatic participants. In the study by Fritz et al[30], acceptable reliability was also reached. Inter-examiner reliability of measurements of passive physiological range of motion ranged from kappa -0.02, for measuring extension before standardisation training[28], to ICC 0.97, for physiotherapists using vision to measure flexion in symptomatic participants[30]. Measuring physiological range of flexion in supine with the hip in 90° flexion consistently yielded acceptable reliability regardless of the method of measurement. Inter-examiner reliability (kappa) of judgements of physiological end-feel ranged from poor (-0.01, 95% CI -0.36 to 0.35), for flexion, to moderate (0.43, 95% CI -0.06 to 0.92), for extension.[31] Both studies[25;31] investigating reliability of end-feel judgements used symptomatic participants.

Table 4. Inter-examiner reliability (95% CI) for judgements of passive physiological end-feel by joint and movement direction

End-feel	Study	Inter-examiner reliability (kappa)
Hip		
Flexion	Currier[25]	0.41 (0.14 to 0.68)
	Sutlive[26]	0.21 (-0.22 to 0.64)
Extension	Currier[25]	-0.13 (-0.48 to 0.22)
Internal rotation	Currier[25]	0.20 (-0.07 to 0.47)
	Sutlive[26]	0.51 (0.19 to 0.83)
External rotation	Currier[25]	-0.02 (-0.37 to 0.33)
Abduction	Currier[25]	0.15 (-0.14 to 0.44)
Adduction	Currier[25]	0.00 (-0.39 to 0.39)
Patrick's test	Currier[25]	0.39 (0.12 to 0.66)
	Sutlive[26]	0.47 (0.12 to 0.81)
Distraction	Currier[25]	0.13 (-0.24 to 0.50)
Scour test	Sutlive[26]	0.52 (0.08 to 0.96)
Knee		
Flexion	Currier[25]	0.31 (-0.53 to 1.00)
	Hayes[31]	-0.01 (-0.36 to 0.35)
Extension	Currier[25]	0.25 (-0.18 to 0.68)
	Hayes[31]	0.43 (-0.06 to 0.92)

Ankle-foot-toes (n = 5)

One study[37] fulfilled all criteria for external validity. In this study, unacceptable inter-examiner reliability was demonstrated by physiotherapists using a goniometer to measure passive physiological range of ankle inversion (ICC 0.42) and eversion (0.25) in symptomatic participants.[37] In the study by Diamond et al[34], acceptable estimates of reliability were reached for measurements of physiological range of ankle dorsiflexion, inversion, and eversion in diabetic patients by well-trained physiotherapists using a goniometer. These estimates could have been underestimated due to instability of characteristics of examiners. Inter-examiner reliability (ICC) of measurements of passive physiological range of motion ranged from 0.00, for measuring ankle dorsiflexion in patients with orthopaedic disorders by trained physiotherapists using a goniometer[35], to 0.88, for measuring ankle inversion[34]. Inter-examiner reliability of measurements of physiological range of motion of the first ray in non-symptomatic participants by podiatric physicians using a goniometer was unacceptable.[27] Finally, the only study[36] in this review investigating accessory range of motion showed fair (kappa 0.35) to moderate (0.48) inter-examiner reliability for measurements of medio-lateral talar motion by physiotherapists in symptomatic participants.

Discussion

This systematic review included 17 studies investigating inter-examiner reliability of passive movements in lower extremity joints. Five studies demonstrated acceptable reliability. In four of these, physiotherapists acted as examiners. Reliability of measurements of physiological range of motion ranged from kappa -0.02, for rheumatologists using a goniometer to measure knee extension in patients with knee osteoarthritis, to ICC 0.97, for physiotherapists using vision to measure knee flexion in symptomatic participants.[28;30] Measuring physiological range of knee flexion consistently yielded acceptable reliability using either vision or instruments. Judgements of end-feel were unreliable for all hip and knee movements. Two high-quality studies reported acceptable reliability for measuring physiological range of knee flexion and extension.[28;33] Overall, however, methodological quality of the included studies was poor.

Inter-examiner reliability for measurement of passive physiological range of motion in lower extremity joints was, overall, considerably less than that in upper extremity joints.[15] In upper extremity joints, measuring large physiological ranges of motion like those in the shoulder, wrist, or fingers using instruments frequently yielded satisfactory reliability.[15] This finding could only partly be confirmed for the lower extremity. For instance, measurement of physiological knee flexion using either vision or instruments indeed showed acceptable reliability, but measurements of relatively smaller ankle movements were unreliable in four out of five studies. However, inter-examiner reliability for hip measurements varied widely across movements and methods of measurement. This heterogeneity in reliability could be explained by the large variation among studies in operational definitions of measurement procedures particularly with respect to participant positioning and instruction, and examiners' execution of movements and handling of instruments. New research investigating inter-examiner reliability for measurement of passive physiological hip movements should incorporate measurement procedures that are in accordance with international standards such as described by Clarkson.[38]

Based on the evidence of three studies[25;26;31], we concluded that judgements of end-feel were unreliable for all hip and knee movements. This conclusion is similar to findings for other regions such as the shoulder, the elbow, and the spinal joints.[8;14;15] Cyriax[2] originally described the concept of end-feel as the different sensations imparted to the hand of the examiner at the extreme of the possible range of joint motion and he believed these were of great diagnostic relevance. This concept has then since long been incorporated in the various international approaches in manual therapy and subsequent educational programs.[39] As a consequence, manual therapists frequently use end-feel as an important indicator of spinal and extremity joint dysfunction.[40-42] The frequency of using end-feel judgements by physiotherapists for diagnosing lower extremity disorders is unknown but assumed to be high. Studies addressing the intra- and inter-examiner reliability of end-feel judgements for diagnosing extremity disorders are needed, with clear and uniform criteria for classifying end-feel.

Only one of the included studies fulfilled all criteria for external validity implying that its results are generalisable to clinical practice.[37] In particular, the majority of studies did not sufficiently describe whether measurements of passive movements were performed with or without clinical information from participants available to examiners. In accordance with guidelines for the methodological quality assessment of diagnostic accuracy studies, we rated Criterion 4 in our quality assessment list (Box 2) as positive when this information would also be available in clinical practice.[18] Presumably, measurements of passive movements of lower extremity joints usually take place after taking a history and performing one or more physical test procedures such as inspection, palpation, resistance tests, provocation tests, or measurement of active movements. Interpretation of measurements of passive movements will then inevitably be influenced by the previously gathered data. This dependence of test results on other information will alter estimates of inter-examiner reliability as opposed to the ones generated by blinded single-test research. In medical test reading, providing clinical information was shown to increase diagnostic accuracy, i.e., sensitivity.[43] Research into the inter-examiner reliability of measurements of passive movements of the extremities should therefore closely resemble clinical practice. However, no data are available on how and when physiotherapists use measurements of passive movements in relation to other diagnostic procedures within their clinical reasoning and decision-making. Identifying the role and position of a test within a diagnostic strategy can help to design studies to evaluate the diagnostic value of tests.[44] In diagnostic research, a stepwise evaluation of tests is increasingly proposed considering not only the test's technical reliability and accuracy but also its place in the clinical pathway and, eventually, its impact on patient outcomes.[45] Investigating the role and position of measurements of passive movements of the extremities within clinical pathways for diagnosing disorders forms an unexplored field of research in physiotherapy and could improve the external validity of future reliability studies.

With respect to internal validity, only two studies satisfied all three criteria suggesting unbiased estimates of inter-examiner reliability.[28;33] This disappointing finding is similar to those of reviews of measurements of upper extremity movements and spinal movement.[11;14;15] However, in many cases, these validity criteria could not be scored due to inadequate reporting of the study protocol. In these cases, it was not possible to provide any indication of the presence and/or direction of the risk of bias. The criteria related to the stability of test circumstances, for both participants and examiners, indicate underestimation of reliability if they are not met. Instability of the participants' characteristics under study – in this case the joint's mobility – may be caused by changes in the biomechanical properties of joint connective tissues as a result of natural variation over time or mobilising effects of the assessment procedure itself.[46] Similarly, instability of the examiners' capability of making judgements may be the result of, for example, mental fatigue. A lack of appropriate blinding of examiners, on the other hand, could lead to overestimation of reliability. If several of these methodological flaws are present, the direction of risk of bias is difficult to predict. Researchers should give careful consideration to ensuring stability of participants' and examiners' characteristics during research and to provide detailed information on the study protocol by following the STARD statement.[16;17] Similar recommendations for improving the reporting of reliability studies were made in the field of medical research.[47]

A lack of inter-examiner reliability adversely affects the accuracy of diagnostic decisions and subsequent treatment selection.[48] This is particularly problematic when effective treatments are available and certain patients run the risk of not receiving them due to error and variation in decision-making among therapists. For instance, hip osteoarthritis is usually defined according to the clinical criteria of the American College of Rheumatology which include criteria about restrictions of physiological range of hip flexion and internal rotation.[49] Hoeksma et al[50] found a beneficial effect of specific manual manipulation and mobilisation of the hip joint on pain, range of motion, and activities in patients with hip osteoarthritis. However, our review did not show acceptable inter-examiner reliability for measuring physiological range of hip flexion and internal rotation. In clinical practice, error and variation in diagnostic classification of hip osteoarthritis may therefore be leaving many patients undertreated. Furthermore, Cyriax's capsular pattern of gross restriction of physiological passive range of hip flexion, abduction, internal rotation and slight restriction of extension for diagnosing hip osteoarthritis was not corroborated making diagnosis based on measurement of passive movements invalid.[2;51;52] Finally, another example in which treatment selection relies on measurement of passive movements is related to the finding that in patients with acute ankle sprain, manual mobilisation or manipulation has an initial beneficial effect on range of ankle dorsiflexion.[53] Only a reliable measurement of restricted ankle dorsiflexion allows a valid decision whether or not to manually intervene. However, measuring passive physiological range of ankle dorsiflexion using a goniometer did not show acceptable reliability. Physiotherapists should incorporate a wider range of findings from their clinical assessment into their decisions about patients with lower extremity disorders and not rely too strongly on results from measurements of passive movements in joints.

Limitations of this study

This review has limitations with respect to its study identification, quality assessment, and data analysis. In our experience, reliability studies were poorly indexed in databases. Although much effort was put in reference tracing and hand searching, eligible studies may have been missed. Furthermore, unpublished studies were not included. Publication bias can threaten the internal validity of systematic reviews of reliability studies because unpublished studies are more likely to report low reliability.

Quality assessment was performed by using a criteria list mainly derived from the assessment of diagnostic accuracy studies. It is not known whether these items also apply in the context of reliability. Empirical evidence of bias, especially concerning blinding of examiners and stability of characteristics of participants and examiners, is lacking. Another method for scoring methodological quality may have resulted in different conclusions. We encourage further validation of the Quality Appraisal of Reliability Studies checklist.[54] Also, study methods were frequently underreported in the included studies. We did not attempt to retrieve more information on study methods from the original authors. Complete information on these methods may have altered our conclusions with respect to study quality.

Finally, our analysis was based on point estimates of reliability. Including interpretation of the precision of these estimates would have provided a more detailed perspective. However, only a limited number of included studies presented 95% CI. In these cases, lower limits never indicated acceptable reliability and most CI were quite wide suggesting low sample sizes. None of the included studies reported an *a priori* sample size calculation.

Conclusions and recommendations

We conclude that the inter-examiner reliability of measurement of passive physiological movements in lower extremity joints is generally low. Future research should focus on determining the role and position of measurements of passive movements in extremity joints within clinical reasoning and decision-making. In addition, the inter-examiner reliability of measurements of passive physiological hip and ankle range of motion in particular and of judgements of end-feel should be further investigated. Careful consideration should be given to uniform standardisation of measurement procedures and to ensuring stability of participants' and examiners' characteristics during research. Sample size calculations should be performed. Finally, following the STARD statement will also improve the quality of reporting of reliability studies. Awaiting new evidence, clinicians should be cautious about relying on results from measurements of passive movements in joints for making decisions about patients with lower extremity disorders.

References

1. Brantingham JW, Globe G, Pollard H, Hicks M, Korporaal C, Hoskins W. Manipulative therapy for lower extremity conditions: Expansion of literature review. J Manipulative Physiol Ther 2009;32:53-71.
2. Cyriax J. Textbook of orthopaedic medicine. Volume one: Diagnosis of soft tissue lesions (8th edition). London, UK: Bailliere Tindall, 1982.
3. Hengeveld E, Banks K. Maitland's peripheral manipulation (4th edition). Philadelphia, PA: Butterworth-Heinemann, 2005.
4. Kaltenborn FM. Manual mobilization of the joints. Volume I: The extremities (6th edition). Oslo, Norway: Olaf Norlis Bokhandel, 2002.
5. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (4th edition) Oxford, UK: Oxford University Press, 2008.
6. De Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
7. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
8. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
9. Hestbæk L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipulative Physiol Ther 2000;23:258-75.
10. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Aust J Physiother 2006;52:91-102.
11. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Spine 2004;29:E413-25.
12. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbaek L, Adams A, Bronfort G. Manual examination of the spine: a systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85.
13. Van der Wurff P, Hagmeijer RH, Meyne W. Clinical tests of the sacroiliac joint. A systematic methodological review. Part 1: Reliability. Man Ther 2000;5:30-6.

14. Van Trijffel E, Anderegg Q, Bossuyt PM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. Man Ther 2005;10:256-69.
15. Van de Pol RJ, Van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. J Physiother 2010;56:7-17.
16. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HCW. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. Clin Chem 2003;49:1-6.
17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, De Vet HCW, Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.
18. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
19. Kramer MS, Feinstein AR. Clinical biostatistics LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981;29:111-23.
20. Landis JR, Koch DG. The measurement of observer agreement for categorical data. Biometrica 1977;33:159–164.
21. Aalto TJ, Airaksinen O, Härkönen TM, Arokoski JP. Effect of passive stretch on reproducibility of hip range of motion measurements. Arch Phys Med Rehabil 2005;86:549-57.
22. Chevillotte CJ, Ali MH, Trousdale RT, Pagnano MW. Variability in hip range of motion on clinical examination. J Arthroplasty 2009;24:693-7.
23. Cibere J, Thorne A, Bellamy N, Greidanus N, Chalmers A, Mahomed N, Shojania K, Kopec J, Esdaile JM. Reliability of the hip examination in osteoarthritis: effect of standardization. Arthritis Rheum 2008;59:373-81.
24. Croft PR, Nahit ES, Macfarlane GJ, Silma AJ. Interobserver reliability in measuring flexion, internal rotation, and external rotation of the hip using a plurimeter. Ann Rheum Dis 1996;55:320-3.
25. Currier LL, Froehlich PJ, Carow SD, McAndrew RK, Cliborne AV, Boyles RE, Mansfield RT, Wainner RS. Development of a clinical prediction rule to identify patients with knee pain

and clinical evidence of knee osteoarthritis who demonstrate a favourable short-term response to hip mobilization. Phys Ther 2007;87:1106-19.

26. Sutlive TG, Lopez HP, Schnitker DE, Yawn SE, Halle RJ, Mansfield LT, Boyles RE, Childs JD. Development of a clinical prediction rule diagnosing hip osteoarthritis in individuals with unilateral hip pain. J Orthop Sports Phys Ther 2008;38:542-50.
27. Van Gheluwe B, Kirby KA, Roosen P, Phillips RD. Reliability and accuracy of biomechanical measurements of the lower extremities. J Am Podiatr Med Assoc 2002;92:317-26.
28. Cibere J, Bellamy N, Thorne A, Esdaile JM, McGorm KJ, Chalmers A, Huang S, Peloso P, Shojania K, Singer J, Wong H, Kopec J. Reliability of the knee examination in osteoarthritis. Arthritis Rheum 2004;50:458-68.
29. Cleffken B, Van Breukelen G, Brink P, Van Mameren H, Olde Damink S. Digital goniometric measurement of knee joint motion. Evaluation of usefulness for research settings and clinical practice. Knee 2007;14:385-9.
30. Fritz JM, Delitto A, Erhard RE, Roman M. An examination of the selective tissue tension scheme, with evidence for the concept of a capsular pattern of the knee. Phys Ther 1998;78:1046-61.
31. Hayes KW, Petersen CM. Reliability of assessing end-feel and pain and resistance sequence in subjects with painful shoulders and knees. J Orthop Sports Phys Ther 2001;31:432-45.
32. Rothstein JM, Miller PJ, Roettger RF. Goniometric reliability in a clinical setting. Elbow and knee measurements. Phys Ther 1983;63:1611-5.
33. Watkins MA, Riddle DL, Lamb RL, Personius WJ. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. Phys Ther 1991;71:90-7.
34. Diamond JE, Mueller MJ, Delitto A, Sinacore DR. Reliability of a diabetic foot evaluation. Phys Ther 1989;69:797-802.
35. Elveru RA, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting. Subtalar and ankle joint measurements. Phys Ther 1988;68:672-7.
36. Erichsen N, Lund H, Møller JO, Kaiser T, Jensen ML, Märcher I, Rune U, Bliddal H. Inter-rater and intra-rater reliability of tests of translatoric movements and range of movements in the subtalar and talocrural joints. Advances in Physiotherapy 2006;8:161-7.

37. Smith-Oricchio K, Harris BA. Interrater reliability of subtalar neutral, calcaneal inversion and eversion. J Orthop Sports Phys Ther 1990;12:10-5.
38. Clarkson HM. Joint motion and function assessment. A research-based practical guide. Philadelphia, PA: Lippincott Williams & Wilkins, 2005.
39. Farrell JP, Jensen GM. Manual therapy: A critical assessment of role in the profession of physical therapy. Phys Ther 1992;72:843-52.
40. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.
41. Van Ravensberg CDD, Oostendorp RAB, Van Berkel LM, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual therapy: Differences in patient characteristics. J Man Manip Ther 2005;13:113-24.
42. Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine. A survey of Dutch physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51.
43. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. A systematic review. Ann Intern Med 2004;140:189-202.
44. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089-92.
45. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F. The evaluation of diagnostic tests: Evidence on technical and diagnostic accuracy, impact on patient outcome and cost effectiveness is needed. J Clin Epidemiol 2007;60:1116-22.
46. Rothstein JM, Echternach JL. Primer on measurement: an introductory guide to measurement issues. Alexandria, VA: American Physical Therapy Association, 1993.
47. Gow RM, Barrowman NJ, Lai L, Moher D. A review of five cardiology journals found that observer variability of measured variables was infrequently reported. J Clin Epidemiol 2008;61:394-401.
48. Quinn MF. Relation of observer agreement to accuracy according to a two-receiver signal detection model of diagnosis. Med Decis Making 1989;9:196-206.
49. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505-14.

50. Hoeksma HL, Dekker J, Ronday HK, Heering A, Van der Lubbe N, Vel C, Breedveld FC, Van den Ende CHM. Comparison of manual therapy and exercise therapy in osteoarthritis of the hip: A randomized clinical trial. Arthritis Rheum 2004;51:722-29.
51. Bijl D, Dekker J, Van Baar ME, Oostendorp RAB, Lemmens AM, Bijlsma JWJ, Voorn ThB. Validity of Cyriax's concept of capsular pattern for the diagnosis of osteoarthritis of hip and/or knee. Scand J Rheumatol 1998;27:347-51.
52. Klässbo M, Harms-Ringdahl K, Larsson G. Examination of passive ROM and capsular patterns in the hip. Physiother Res Int 2003;8:1-12.
53. Van der Wees PJ, Lenssen AF, Hendriks EJM, Stomp DJ, Dekker J, De Bie RA. Effectiveness of exercise therapy and manual mobilisation in acute ankle sprain and functional instability: A systematic review. Aust J Physiother 2006;52:27-37.
54. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010;63:854-61.

Appendix 1

Search strategy for MEDLINE (PubMed) using text words [tw] and Medical Subject Headings [mh]

1. hip [mh]
2. hip [tw]
3. knee [mh]
4. knee [tw]
5. patellofemoral [tw]
6. tibiofibular [tw]
7. ankle [mh]
8. ankle [tw]
9. talocrural [tw]
10. subtalar [tw]
11. talocalcaneal [tw]
12. foot [mh]
13. foot [tw]
14. tarsal [tw]
15. midtarsal [tw]
16. intertarsal [tw]
17. calcaneocuboid [tw]
18. talocalcaneonavicular [tw]
19. cuneometatarsal [tw]
20. cuneonavicular [tw]
21. tarsometatarsal [tw]
22. metatarsophalangeal [tw]
23. phalangeal [tw]
24. interphalangeal [tw]
25. toe* [mh]
26. toe* [tw]

27. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26
28. 27 AND Joint* [mh]
29. motion [mh]
30. movement [mh]
31. range of motion, articular [mh]
32. mobility [tw]
33. endfeel [tw]
34. end feel [tw]
35. 29 OR 30 OR 31 OR 32 OR 33 OR 34
36. physical examination [mh]
37. diagnostic tests, routine [mh]
38. observation [mh]
39. passive [tw]
40. manual [tw]
41. 36 OR 37 OR 38 OR 39 OR 40
42. reproducibility of results [mh]
43. reproducibility [tw]
44. reliability [tw]
45. observer variation [tw]
46. repeatability [tw]
47. variation [tw]
48. concordance [tw]
49. variability [tw]
50. agreement [tw]
51. 42 OR 43 OR 44 OR 45 OR 46 OR 47 OR 48 OR 49 OR 50
52. interobserver [tw]
53. intertester [tw]
54. interrater [tw]
55. interexaminer [tw]
56. observer* [tw]
57. tester* [tw]

58. rater* [tw]

59. examiner* [tw]

60. 52 OR 53 OR 54 OR 55 OR 56 OR 57 OR 58 OR 59

61. 28 AND 35 AND 41 AND 51 AND 60

Appendix 2

Excluded studies *(n = 31)* with their main reason for exclusion

Alexander RE, Battye CK, Goodwill CJ, Walsh JB. The ankle and subtalar joints. Clin Rheum Dis 1982;8:703-11.
Reason for exclusion: No reliability study

Arokoski MH, Haara M, Helminen HJ, Arokoski JP. Physical function in men with and without hip osteoarthritis. Arch Phys Med Rehabil 2004;85:574-81.
Reason for exclusion: No inter-examiner reliability study

Backer M, Kofoed H. Passive ankle mobility. Clinical measurement compared with radiography. J Bone Joint Surg Br 1989;71:696-8.
Reason for exclusion: No reliability study

Ball P, Johnson GR. Reliability of hindfoot goniometry when using a flexible electrogoniometer. Clin Biomech 1993;8:13-9.
Reason for exclusion: Using instruments not feasible in practice

Bierma-Zeinstra SM, Bohnen AM, Ramlal R, Ridderikhoff J, Verhaar JA, Prins A. Comparison between two devices for measuring hip joint motions. Clin Rehabil 1998;12:497-505.
Reason for exclusion: No estimates of inter-examiner reliability reported

Buckley RE, Hunt DV. Reliability of clinical measurement of subtalar movement. Foot Ankle Int 1997;18:229-32.
Reason for exclusion: No estimates of inter-examiner reliability reported

Cliborne AV, Wainner RS, Rhon DI, Judd CD, Fee TT, Matekel RL, Whitman JM. Clinical hip tests and a functional squat test in patients with knee osteoarthritis: reliability, prevalence of positive test findings, and short-term response to hip mobilization. J Orthop Sports Phys Ther 2004;34:676-85.
Reason for exclusion: No inter-examiner reliability study

Cushnaghan J, Cooper C, Dieppe P, Kirwan J, McAlindon T, McCrae F. Clinical assessment of osteoarthritis of the knee. Ann Rheum Dis 1990;49:768-70.
Reason for exclusion: No passive range of motion measurements

Ekstrand J, Wiktorsson M, Oberg B, Gillquist J. Lower extremity goniometric measurements: a study to determine their reliability. Arch Phys Med Rehabil 1982;63:171-5.
Reason for exclusion: No inter-examiner reliability study

Ellison JB, Rose SJ, Sahrmann SA. Patterns of hip rotation range of motion: a comparison between healthy subjects and patients with low back pain. Phys Ther 1990;70:537-41.
Reason for exclusion: No reliability study

Fredriksen H, Dagfinrud H, Jacobsen V, Maehlum S. Passive knee extension test to measure hamstring muscle tightness. Scand J Med Sci Sports 1997;7:279-82.
Reason for exclusion: Using instruments not feasible in practice

Gogia PP, Braatz JH, Rose SJ, Norton BJ. Reliability and validity of goniometric measurements at the knee. Phys Ther 1987;67:192-5.
Reason for exclusion: No passive range of motion measurements

Hayes KW, Petersen C, Falconer J. An examination of Cyriax's passive motion tests with patients having osteoarthritis of the knee. Phys Ther 1994;74:697-707.
Reason for exclusion: No inter-examiner reliability study

Holm I, Bolstad B, Lütken T, Ervik A, Røkkum M, Steen H. Reliability of goniometric measurements and visual estimates of hip ROM in patients with osteoarthrosis. Physiother Res Int 2000;5:241-8.
Reason for exclusion: No estimates of inter-examiner reliability reported

Hopson MM, McPoil TG, Cornwall MW. Motion of the first metatarsophalangeal joint. Reliability and validity of four measurement techniques. J Am Podiatr Med Assoc 1995;85:198-204.
Reason for exclusion: No inter-examiner reliability study

Jonson SR, Gross MT. Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipmen. J Orthop Sports Phys Ther 1997;25:253-63.
Reason for exclusion: No passive range of motion measurements

Kim JY, Keun Hwang S, Tai Lee K, Won Young K, Seon Jung J. A simpler device for measuring the mobility of the first ray of the foot. Foot Ankle Int 2008;29:213-8.
Reason for exclusion: No passive range of motion measurements

Klässbo M, Harms-Ringdahl K, Larsson G. Examination of passive ROM and capsular patterns in the hip. Physiother Res Int 2003;8:1-12.
Reason for exclusion: No inter-examiner reliability study

Macedo LG, Magee DJ. Differences in range of motion between dominant and nondominant sides of upper and lower extremities. J Manipulative Physiol Ther 2008;31:577-82.
Reason for exclusion: No inter-examiner reliability study

Marks JS, Palmer MK, Burke MJ, Smith P. Observer variation in examination of knee joints. Ann Rheum Dis 1978;37:376-7.
Reason for exclusion: No inter-examiner reliability study

Meyer DC, Werner CM, Wyss T, Vienne P. A mechanical equinometer to measure the range of motion of the ankle joint: interobserver and intraobserver reliability. Foot Ankle Int 2006;27:202-5.
Reason for exclusion: Using instruments not feasible in practice

Moseley A, Adams R. Measurement of passive ankle dorsiflexion: procedure and reliability. Aust J Physiother 1991;37:175-12.
Reason for exclusion: Using instruments not feasible in practice

Oberg U, Oberg B, Oberg T. Validity and reliability of a new assessment of lower-extremity dysfunction. Phys Ther 1994;74:861-71.
Reason for exclusion: No passive range of motion measurements

Peeler J, Anderson JE. Reliability of the Ely's test for assessing rectus femoris muscle flexibility and joint range of motion. J Orthop Res 2008;26:793-9.
Reason for exclusion: No passive range of motion measurements

Petersen CM, Hayes KW. Construct validity of Cyriax's selective tension examination: association of end-feels with pain at the knee and shoulder. J Orthop Sports Phys Ther 2000;30:512-21.
Reason for exclusion: Duplicate publication

Piva SR, Fitzgerald K, Irrgang JJ, Jones S, Hando BR, Browder DA, Childs JD. Reliability of measures of impairments associated with patellofemoral pain syndrome. BMC Musculoskelet Disord 2006;7:33.
Reason for exclusion: No passive range of motion measurements

Ross MD, Nordeen MH, Barido M. Test-retest reliability of Patrick's hip range of motion test in healthy college-aged men. J Strength Cond Res 2003;17:156-61.
Reason for exclusion: No inter-examiner reliability study

Salsich GB, Mueller MJ, Sahrmann SA. Passive ankle stiffness in subjects with diabetes and peripheral neuropathy versus an age-matched comparison group. Phys Ther 2000;80:352-62.
Reason for exclusion: No reliability study

Theiler R, Stucki G, Schütz R, Hofer H, Seifert B, Tyndall A, Michel BA. Parametric and non-parametric measures in the assessment of knee and hip osteoarthritis: interobserver reliability and correlation with radiology. Osteoarthritis Cartilage 1996;4:35-42.
Reason for exclusion: Unknown whether active or passive movements were used

Vtasalo JT, Kvist M. Some biomechanical aspects of the foot and ankle in athletes with and without shin splints. Am J Sports Med 1983;11:125-30.
Reason for exclusion: No reliability study

Weiner DK, Sakamoto S, Perera S, Breuer P. Chronic low back pain in older adults: prevalence, reliability, and validity of physical examination findings. J Am Geriatr Soc 2006;54:11-20.
Reason for exclusion: No passive range of motion measurements

Chapter 4

Passive motion assessment of the spine in clinical practice in manual therapy

Chapter 4a

Perceptions and use of passive intervertebral motion assessment of the spine: A survey among manual therapists

Emiel van Trijffel, Rob Oostendorp, Robert Lindeboom, Patrick Bossuyt, Cees Lucas

Man Ther 2009;14:243-51

Abstract

Background: Manual therapists commonly use passive intervertebral motion (PIVM) assessment within physical examination. Data describing the use and interpretation of this manual diagnostic procedure, as well as therapists' perception of related importance and confidence, are lacking.

Methods: A survey was conducted among Dutch manual therapists using a 13-item, self-administered, structured questionnaire exploring demographic and professional characteristics, the use of PIVM assessment, and perceived importance and confidence related to PIVM assessment.

Results: Three hundred and sixty-seven questionnaires were analysed. Response rate from the postal part of the survey was 56%. Dutch manual therapists most frequently apply PIVM assessment to the cervical region and they prefer three-dimensionally coupled motions. They consider end-feel or, to a lesser extent, provocation of patient's pain as decisive for diagnostic conclusions. Respondents believe that these spinal motion tests are important for treatment decisions and are confident in their conclusions drawn from it. These perceptions were largely stable across subgroups of therapists with different gender, age, experience, and educational background. Weekly amount of work related to spinal disorders was positively associated with perceived importance and confidence.

Conclusions: Reported use and interpretation of PIVM assessment and related perceptions could only partly be substantiated by evidence. Results from this survey will help researchers design studies better reflecting clinical practice in manual therapy.

Introduction

The Dutch Association for Manual Therapy describes manual therapy as a specialisation within physiotherapy characterised by the analysis, interpretation, and treatment of complex health problems resulting from arthrogenic, muscular, and neurogenic disorders of the spinal column and extremities using specific manual diagnostic and manual therapeutic techniques.[1] Contrary to many other countries, in the Netherlands, manual therapy is considered a post-graduate (non-university) specialisation within physiotherapy providing practitioners additional knowledge and skills for manual diagnosis and high-velocity thrust interventions.[2;3] Dutch physiotherapists specialising in manual therapy (manual therapists) have explicitly been profiling themselves as specialists in the care of health problems arising from spine-related disorders.[4]

Manual therapy is characterised by the skill of therapists to induce articulatory movements manually in joints of spinal motion segments like, for instance, passive physiological and accessory movements.[1;5-7] From a diagnostic perspective, judging the quantity and quality of passive segmental intervertebral joint motion contributes to the classification of patients.[8]

Little is known about how manual therapists and physiotherapists use and interpret passive intervertebral motion (PIVM) assessment within clinical decision-making. Dutch manual therapists significantly more often detected impairments of joint mobility than Dutch physiotherapists did.[3;7] Manual therapists participating in these studies believed ‘joint range of motion’ and ‘manual end-feel’ are relevant indicators of such impairments. A survey among orthopaedic certified specialists from the American Physical Therapy Association revealed that ‘segmental mobility testing or pain provocation’ was often used for the diagnosis of clinical lumbar instability.[9] Australian physiotherapists rated the presence of an ‘excessively free end-feel’ on passive motion testing as highly important in the recognition of minor cervical instability.[10] However, it remains unclear how manual therapists use, judge, and interpret PIVM assessment within their diagnostic reasoning leading to treatment decisions. In addition, it is unknown to what extent they believe this diagnostic procedure is important for decision-making or how confident they are in their conclusions drawn from it.

A cross-sectional study using a self-administered survey questionnaire was conducted to describe and explore the use of PIVM assessment by Dutch manual therapists and, additionally, to identify factors associated with therapists' perception of related importance and confidence.

Methods

Survey instrument

We developed a 13-item, structured questionnaire aimed at exploring the following three domains: demographic and professional characteristics, the use of PIVM assessment, and perceived importance and confidence related to PIVM assessment (Table 1).

Table 1. Survey instrument consisting of 13 items divided into three domains

Domain	Items
Demographic and professional characteristics	Gender
	Age
	Weekly amount of work related to spinal disorders
	MT educational background
	Experience in MT
Use of PIVM assessment in clinical practice	Most frequently examined spinal region
	Most frequently applied type of movement
	Most decisive clinical finding
	Scale(s) used for categorising clinical findings
	Term(s) used for recording of identified impairments of function of motion segments
Perceived importance and confidence related to PIVM assessment	Importance of PIVM assessment for treatment decisions
	Confidence in reaching correct diagnostic conclusions with PIVM assessment
	Confidence in reaching the same diagnostic conclusions with PIVM assessment as compared to a random colleague

MT: manual therapy, PIVM: passive intervertebral motion

In the second domain, two open-ended questions were used inviting respondents to describe types of scales used for classifying clinical findings and terms used for recording of identified impairments of function of vertebral motion segments in patient records. See Appendix for definitions of types of movements applied for PIVM assessment.[11-14] In the third domain, respondents rated their perceived importance and confidence on a 7-point rating scale.

Procedure

The questionnaire was tested for interpretability in two groups of manual therapists constituting consultation platforms. These platforms are part of the quality assurance program of the Royal Dutch Society for Physical Therapy and generally consist of up to 15 therapists working on quality improvement and assurance.[15] These testing rounds led to minor rephrasing of two items. Completing the questionnaire took 3-5 min.

The final version of the questionnaire was sent by email to all practices in the Netherlands listed under 'Manual therapists' in the Yellow Pages[16] and the Telephone Guide[17] databases with a link to their email address (September 2006). Potential respondents were requested to complete the questionnaire and return it by email within three weeks. Consequently, from a single practice, more than one manual therapist could potentially respond. A reminder, accompanied by a new copy of the questionnaire, was sent after one month. Next, questionnaires were sent by post to all 23 manual therapy consultation platforms in the Netherlands (November 2006). Members were asked to complete the questionnaires during their next meeting and return these using a prepaid and pre-addressed envelope. After two months, a reminder was sent in which the opportunity was given to request for new copies of the questionnaire. Finally, a random selection of 200 manual therapists out of 2796 (as at January 1, 2007) registered in the Quality Register of the Royal Dutch Society for Physical Therapy[18] received a copy of the questionnaire by post (February 2007). Simultaneously, a random sample of 200 practices for manual therapy listed in the Telephone Guide database[19] also received one questionnaire each by post. Practices involved in the email survey were excluded. Wherever possible, personal addressing was used. Respondents were asked to complete and return the questionnaire within three weeks using a prepaid and pre-addressed envelope. No reminder was sent.

We incorporated methods that have been proven to increase response rates to postal questionnaires.[20] Potential respondents were informed by means of a cover letter explaining the purpose of the study. In case of multiple choice items, they were explicitly requested to select one answer only. It was also pointed out to them that data processing would be carried out anonymously. They were explicitly asked not to return questionnaires twice.

Statistical analysis

Absolute and relative frequencies were used to describe categorical data. Ordinal data relating to the perceived importance and confidence items from the third domain of the questionnaire were additionally described with their medians and interquartile ranges (IQR). Normally distributed numerical data were summarised by their means and standard deviations. In case of non-normal distribution, median and range were presented. Answers to the two open-ended questions in the second domain were recorded and ranked according to reported frequency. Internal consistency reliability of the domain containing the importance and confidence items relating to conclusions drawn from PIVM assessment was calculated using Cronbach's alpha. An alpha >0.70 indicates homogeneity of the domain and consistency in scoring among respondents.[21] Rasch rating scale analysis was used to examine the reliability of the rating scale structure of the importance and confidence items using an item response theory measurement model and OPLM, a computer software program for Rasch measurement models.[22] Guided by this rating scale analysis, rating categories were dichotomised to obtain the best discrimination between respondents' perceptions. Subsequently, univariable logistic regression was performed to identify demographic and professional characteristics of respondents that were associated with perceived importance and confidence. Strength of associations was expressed as odds ratios (OR) with their 95% CI and corresponding p-values. An OR of 1 indicates no association between the importance or confidence item and the demographic or professional characteristic while an OR much greater or less than 1 indicates stronger associations. All analyses were carried out in SPSS (version 14.0). Missing data were not replaced. Multiple answers to multiple choice questions were handled as missing.

Results

Response rates

From 858 citations found in the Yellow Pages and 1079 found in the Telephone Guide, 178 and 128 practices, respectively, had a link to their email address. Twenty-eight practices responded within three weeks and returned 33 questionnaires. After the reminder, another 21 questionnaires (21 practices) were received bringing the email response rate to 16% (49/306). Ten consultation platforms responded by returning 68 questionnaires. After the reminder, another 22 questionnaires were sent by two platforms yielding a response rate of 52% (12/23). Finally, 223 (56%) completed questionnaires were returned by post. In total, 367 questionnaires, containing 31 (0.7%) missing data, were analysed.

Descriptive findings

Demographic and professional characteristics of the survey sample are summarised in Table 2.

Two hundred and seventy-four respondents (76.8%) reported applying PIVM assessment most frequently to the cervical region, i.e., motion segments C0-T4 (Figure).

When using segmental motion assessment, almost 80% (291/366) of manual therapists most frequently apply three-dimensional (coupled) physiological motions while 23 (6.3%) indicated they use one-dimensional physiological movements primarily and about 11% (39/366) prefer accessory motion assessment.

Forty-eight percent of respondents (176/367) consider perceived resistance at the end of the movement (end-feel) as the most decisive clinical finding from PIVM assessment for making diagnostic conclusions about impairments of joint function of motion segments while 22.6% (83/367) preferred provocation or reduction of pain or other symptoms for this purpose. Forty-eight (13.1%) therapists reported to judge PIVM primarily on range of motion and 10.4% (38/367) relied on perceived resistance during movement.

Ninety-two manual therapists (25.1%) stated they made explicit use of scales for categorising clinical findings from PIVM assessment. For classifying end-feel (19 times), scales were used with terms like ‘hard’, ‘empty’, ‘springy’, and ‘stiff’ most often reported. Visual analogue scales (22 times) were the scale of choice for measuring patient’s pain. Nine

Table 2. Demographic and professional characteristics of the survey sample *(n = 367)*

Male gender	281 (86.6%)
	missing 0
Mean age (SD)	46.1 (8.0) yr
	missing 0
Weekly amount of work related to spinal disorders*	24.0 (1.0 – 55.0) hr
	missing 6
MT educational background	
SOMT	241 (66.8%)
MT Utrecht (Van der Bijl)	39 (10.8%)
Maitland's Concept	19 (5.3%)
Vrije Universiteit Brussel Master MT	3 (0.8%)
Orthopaedic MT	31 (8.5%)
Other	5 (1.4%)
More than one	23 (6.4%)
	missing 6
Experience in MT*	14.0 (1.0 – 40.0) yr
	missing 7

MT: manual therapy, SD: standard deviation, SOMT: Stichting Opleidingen Musculoskeletale Therapie (Institute for Master Education in Musculoskeletal Therapy)
*Data described as median (minimum-maximum)

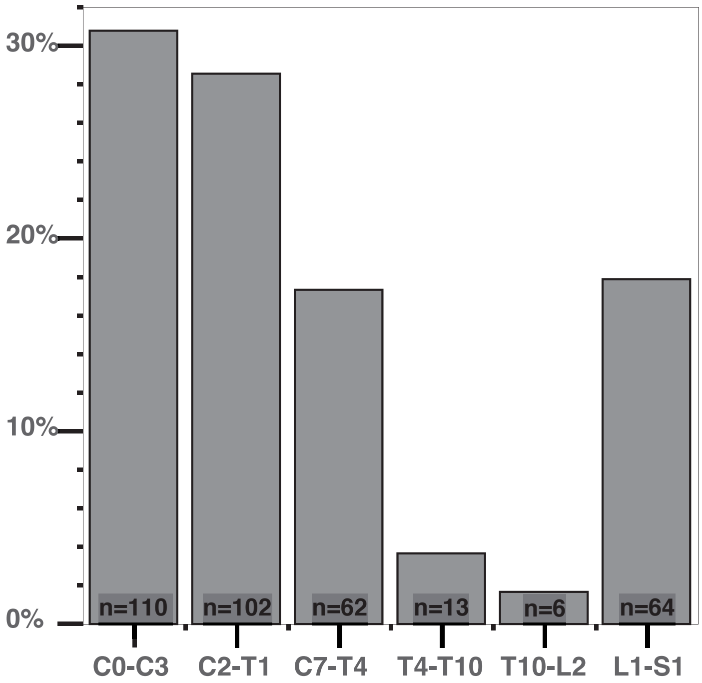

Figure. Bar chart showing absolute (in bars) and relative (y-axis) frequencies of most frequently examined spinal region using passive intervertebral motion assessment assessment (n=357)

therapists reported their use of Maitland's movement diagram for grading mobility and a three-point scale (hypomobile-normal-hypermobile) for this purpose was mentioned seven times.

In total, 67 different terms were given for the recording of identified impairments of motion segments in patient records. Table 3 shows the 10 most frequently reported terms. Some respondents additionally recorded segmental level (35 times) or motion direction (36) of impairments, or both (24).

In Table 4, frequencies of scores on the importance and confidence items are presented. Eighty-one percent (296/367) of respondents believed that diagnostic conclusions from PIVM assessment were reasonably or very important for deciding on manual therapy as a treatment option (IQR 'reasonably important' to 'very important'). With respect to perceived confidence in diagnostic conclusions drawn from PIVM assessment, 198 therapists (54.0%)

Table 3. Ranked, absolute frequencies of the 10 most frequently reported terms for recording of identified impairments of motion segments in patient records *(n = 367)*

Block	77
Restriction	47
Motion restriction	38
Restricted	34
Functional impairment	28
Hypomobility	19
Hypermobility	17
Instability	13
Hypofunction	8
Dysfunction	6

were reasonably confident that they would reach a correct diagnosis about impairments of function of motion segments (IQR 'somewhat confident' to 'reasonably confident') while 251 (68.4%) were somewhat or reasonably confident that they would reach the same conclusions as a random colleague (IQR 'neutral' to 'reasonably confident'). Cronbach's alpha for the total domain was 0.75, indicating that, on the whole, respondents were consistent in their reporting of perceptions.

Inferential findings

Rating scale analysis indicated that collapsing the 'reasonably important' and 'very important' and the 'reasonably confident' and 'very confident' categories, versus the collapsed remaining five categories, offered the best differentiation between respondents' scores on perceived importance and confidence regarding the use of PIVM assessment. ORs representing strengths of associations between the three recoded dichotomous variables on the one hand and demographic and professional characteristics on the other are shown in Table 5.

Weekly amount of work related to spinal disorders was positively associated with all perceptions of importance and confidence. This means, for example, that for every additional weekly hour spent on treating patients with health problems arising from

Table 4. Frequencies of scores on perceived importance and confidence related to passive intervertebral motion assessment ***(n = 367)***

How important to you are diagnostic conclusions from PIVM assessment for deciding on manual therapy as a treatment option?						
Very unimportant	Reasonably unimportant	Somewhat unimportant	Neutral	Somewhat important	Reasonably important	Very important
5 (1.4%)	3 (0.8%)	2 (0.5%)	5 (1.4%)	56 (15.3%)	198* (53.9%)	98 (26.7%)

How confident are you by using PIVM assessment in reaching the correct diagnostic conclusions with regard to impairments of motion segments?						
Very unconfident	Reasonably unconfident	Somewhat unconfident	Neutral	Somewhat confident	Reasonably confident	Very confident
5 (1.4%)	10 (2.7%)	16 (4.4%)	29 (7.9%)	99 (26.9%)	198* (54.0%)	10 (2.7%)

How confident are you by using PIVM assessment in reaching the same diagnostic conclusions as a random colleague with the same educational background?						
Very unconfident	Reasonably unconfident	Somewhat unconfident	Neutral	Somewhat confident	Reasonably confident	Very confident
11 (3.0%)	20 (5.4%)	31 (8.4%)	45 (12.3%)	113* (30.8%)	138 (37.6%)	9 (2.5%)

PIVM: passive intervertebral motion
*median score

disorders of the vertebral column, there was a three percent higher chance (odds) to believe diagnostic conclusions from PIVM assessment are ‘reasonably important’ or ‘very important’ for treatment decisions. Similarly, therapists trained according to the orthopaedic manual therapy principles and Maitland’s Concept were more than twice and three times, respectively, more likely to be ‘reasonably confident’ or ‘very confident’ in reaching the same diagnostic conclusions as their colleagues as compared to respondents educated by the Institute for Master Education in Musculoskeletal Therapy (SOMT: Stichting Opleidingen Musculoskeletale Therapie).

Table 5. Univariable logistic regression analysis using scores on importance and confidence regarding PIVM assessment as dependent factors and demographic and professional characteristics as independent explanatory variables *(n = 367)*

	Importance of PIVM assessment for treatment decisions			Confidence in reaching correct diagnostic conclusions with PIVM assessment			Confidence in reaching the same diagnostic conclusions with PIVM assessment compared to a colleague		
Characteristic	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Male gender	1.31	[0.69,2.49]	0.412	0.90	[0.55,1.46]	0.665	0.91	[0.56,1.50]	0.716
Age	0.98	[0.95,1.01]	0.205	1.03	[0.99,1.05]	0.067	1.00	[0.97,1.02]	0.863
Weekly amount of work related to spinal disorders	1.03	[1.01,1.06]	0.033*	1.04	[1.02,1.07]	<0.0001*	1.03	[1.01,1.06]	0.007*
MT educational background#									
Maitland's Concept	1.03	[0.29,3.70]	0.964	1.24	[0.47,3.25]	0.667	3.03	[1.15,7.99]	0.025*
Orthopaedic MT	0.66	[0.27,1.64]	0.374	1.31	[0.60,2.86]	0.495	2.15	[1.01,4.57]	0.047*
Experience in MT	0.98	[0.95,1.01]	0.197	1.02	[0.99,1.05]	0.180	1.00	[0.97,1.03]	0.877

MT: manual therapy, OR: odds ratio, PIVM: passive intervertebral motion, SOMT: Stichting Opleidingen Musculoskeletale Therapie (Institute for Master Education in Musculoskeletal Therapy)

*Significant at the 0.05 level, #Reference category: SOMT (results from MT Utrecht (Van der Bijl) and Vrije Universiteit Brussel Master MT not shown)

Discussion

This survey found that Dutch manual therapists most frequently apply PIVM assessment to the cervical region and prefer three-dimensional coupled motions. They consider end-feel or, to a lesser extent, pain or other symptoms and range of motion as the decisive clinical finding for diagnostic conclusions concerning impairments of motion segments. Practitioners believe that this manual diagnostic procedure is important for deciding on manual therapy as a treatment option and they are confident in their conclusions drawn from it. These reported perceptions were largely stable across subgroups of therapists with different gender, age, experience, and educational background.

The majority of respondents reported applying PIVM testing most often to the cervical region with about 31% choosing the upper cervical spine in particular. A systematic review showed acceptable inter-examiner reliability of passive assessment of motion in segments C1-C2 and C2-C3.[23] This has been confirmed for assessment of rotational mobility of C1-C2 in later studies.[24;25] Another study showed a high level of reliability for lateral gliding examination of C0-C1.[26] However, findings from PIVM assessment were not included in clinical prediction rules for guiding manipulative treatment of patients with neck pain.[27;28] Segmental hypomobility of lumbar motion segments, on the other hand, has been recognised within a validated prediction rule as a predictor of a successful outcome after spinal manipulation in patients with low-back pain.[29;30] The lumbar spinal column was reported by only 20% of our sample as the region most frequently examined.

Dutch manual therapists prefer to use three-dimensionally coupled movements for passive segmental motion assessment. Cramer et al[31] concluded that, although all spinal motions are indeed coupled motions, motion patterns are complex and coupling differs from one segment to the other. Coupling behaviour of the lumbar and thoracic spine has been shown to be inconsistent with respect to directions in which side-bending and axial rotations are associated.[12;32;33] With respect to the cervical spine, there is full agreement about coupling behaviour of motion segments C2-T1 but variation exists in patterns of C0-C1 and C1-C2.[14] Inter-examiner reliability of passive three-dimensional movement tests was poor for motion segment L4-L5 while in the mid-thoracic spine (T6-T7) fair to substantial agreement beyond chance was obtained.[13;34] Because of all these variations and the unpredictability of coupling biomechanics in pathological states, authors have cautioned to be reticent in the use of three-dimensional motion assessment in patients.[35;36]

We found considerable variation among the sample with respect to which clinical finding from PIVM assessment would be decisive for diagnostic conclusions about impairments of function of motion segments. Jull et al[8] proposed to guide detection of dysfunctional spinal segments by assessing tissue stiffness via the presence of muscle reactivity or abnormal thicker through range resistance. In Maitland's Concept, change in resistance perceived by the therapist during movement combined with pain reported by the patient is used to construct movement diagrams.[37] Only a small proportion of respondents chose resistance perceived during passive motion testing as an important diagnostic phenomenon. A recent survey among manual physical therapists in New Zealand and the USA revealed that passive accessory lumbar segmental motion testing is performed to assess pain response and quality of resistance, and physiological motion testing is used to assess quality of motion path.[38] No consensus exists on which clinical finding - or combination of clinical findings - is appropriate to identify impairments of function of motion segments nor on which scale to use for categorising findings. Likewise, 67 different terms were identified for the recording of impairments of motion segments. It seems that manual therapists suffer from the same lack of uniformity in terminology as their colleagues in chiropractic do.[39]

Dutch practitioners are confident to be correct in their conclusions drawn from PIVM assessment. Reaching correct diagnoses about impairments of function of motion segments reflects the validity – or diagnostic accuracy – of the test procedure. Evidence of accuracy of segmental motion testing is accumulating gradually but it does not permit definitive conclusions.[25;40-44] Respondents are somewhat less confident in reaching the same diagnostic conclusions from PIVM assessment as compared to a random colleague. This means that they are less confident in the inter-examiner reliability of passive segmental motion testing than in its diagnostic accuracy. Nevertheless, reported levels of confidence in reliability were not in accordance with available evidence. Inter-examiner reliability of segmental intervertebral motion tests has been found to be unacceptably low.[23;45-47] Seffinger et al[45] concluded that assessing regional range of spinal motion was more reliable than segmental examination. Several authors have questioned the clinical usefulness and necessity of identifying impairments of joint mobility at specified spinal levels in order to make treatment decisions.[11;48-50]

Limitations of this study

Low response rates in survey research reduce sample size and precision as well as threaten the validity in case non-responders may differ systematically from responders.[51] Among health professionals, response rates to mail surveys vary widely, from 16% to 91%.[52] The response rate from the postal part of our survey was comparable to rates among other care providers.[53;54] Response to the email component, on the contrary, was at the very low end of the range. We did not collect data on non-responders and data on the distribution of characteristics of manual therapists registered by the Dutch Association for Manual Therapy or the Royal Dutch Society for Physical Therapy were not available. It might be possible that non-responders would have scored systematically different with respect to their use and perceptions of PIVM assessment which could have biased our results. Given the large sample size achieved, we assume distribution of type of educational background in our survey sample to be correctly reflecting the total population of Dutch manual therapists.

Furthermore, manual therapy education in the Netherlands is strongly embedded within international concepts. In these traditional concepts, passive joint motion assessment has been provided a prominent place.[5] Therefore, we suppose that results of this study will to a certain extent be generalisable to populations of manual therapists outside the Netherlands. Our opinion is partly supported by Abbott et al[38] showing that manual physical therapists from New Zealand and the USA believe passive accessory and physiological motion testing is accurate for estimating the quantity of movement present at a lumbar segment and segmental motion findings are important for treatment selection.

Finally, respondents could potentially have returned more than one questionnaire each. Because priority was given to anonymous data processing, we were unable to control this possible threat to validity in our results.

Conclusions and recommendations

Dutch manual therapists showed substantial consistency in reporting their use, interpretation, and related perceptions of importance and confidence regarding PIVM assessment. However, these findings could only partly be substantiated by evidence. The role and position of PIVM testing of the spine within the diagnostic pathway as a whole need further clarification to allow more useful evaluation of its diagnostic value.[55] We aim that the results of this survey will guide future research to better reflect clinical practice in manual therapy.

References

1. Dutch Association for Manual Therapy. Professional competence profile for the manual therapist (May 2005). Amersfoort, The Netherlands. Available at: http://www.nvmt.nl/upload/BCP.ENG.pdf. (accessed 2 February, 2007)
2. Oostendorp RAB, Scholten-Peeters GGM, Swinkels RAHM, Bekkering GE, Heijmans MWFGJ, Huijbregts PA, Hendriks EJM. Evidence-based practice in physical and manual therapy: Development and content of Dutch national practice guidelines for patients with non-specific low back pain. J Man Manip Ther 2004;12:21-31.
3. Oostendorp RAB, Van Berkel LM, Van Ravensberg CDD, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual therapy for patients with non-specific low-back pain: Differences in patient characteristics with implications for diagnostic classification. J Man Manip Ther 2006;14:E46-55.
4. Dutch Association for Manual Therapy. Vision document (June 2001). Amersfoort, The Netherlands. Available at: http://www.nvmt.nl/upload/Visiedocument.pdf. [In Dutch] (accessed 9 March, 2007)
5. Farrell JP, Jensen GM. Manual therapy: A critical assessment of role in the profession of physical therapy. Phys Ther 1992;72:843-52.
6. Maher C, Latimer J. Pain or resistance – the manual therapists' dilemma. Aust J Physiother 1992;38:257-60.
7. Van Ravensberg CDD, Oostendorp RAB, Van Berkel LM, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual therapy: Differences in patient characteristics. J Man Manip Ther 2005;13:113-24.
8. Jull G, Treleaven J, Versace G. Manual examination: Is pain provocation a major cue for spinal dysfunction? Aust J Physiother 1994;40:159-65.
9. Cook C, Brismée J-M, Sizer PS. Factors associated with physiotherapists' confidence during assessment of clinical cervical and lumbar spine instability. Physiother Res Int 2005;10:59-71.
10. Niere KR, Torney SK. Clinicians' perceptions of minor cervical instability. Man Ther 2004;9:144-50.
11. Huijbregts PA. Spinal motion palpation: A review of reliability studies. J Man Manip Ther 2002;10:24-39.

12. Cook C. Coupling behaviour of the lumbar spine: A literature review. J Man Manip Ther 2003;11:137-45.
13. Brismée J-M, Gipson D, Ivie D, Lopez A, Moore M, Matthijs O, Phelps V, Sawyer S, Sizer P. Interrater reliability of a passive physiological intervertebral motion test in the mid-thoracic spine. J Manipulative Physiol Ther 2006;29:368-73.
14. Cook C, Hegedus E, Showalter C, Sizer PS. Coupling behaviour of the cervical spine: A systematic review of the literature. J Manipulative Physiol Ther 2006;29:570-5.
15. Van der Wees PJ, Hendriks EJM, Veldhuizen RJ. Quality assurance in the Netherlands: From development to implementation and evaluation. Dutch Journal of Physical Therapy (Special Issue International WCPT Congress Barcelona 2003) 2003;113(S):3-6.
16. Yellow Pages 2006. Available at: http://www.goudengids.nl/search/Manueel_therapeuten.html. [In Dutch] (accessed 1 September, 2006)
17. Telephone Guide 2006. Available at: http://dtgi.detelefoongids.nl. [In Dutch] (accessed 1 September, 2006)
18. Royal Dutch Society for Physical Therapy 2007. Available at: http://www.fysiotherapie.nl/index.html. [In Dutch] (accessed 14 February, 2007)
19. Telephone Guide 2007. Available at: http://dtgi.detelefoongids.nl. [In Dutch] (accessed 31 January, 2007)
20. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, Kwan I, Cooper R. Methods to increase response rates to postal questionnaires. The Cochrane Database of Methodology Reviews 2003, Issue 4. Art. No.: MR000008.
21. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (3rd edition). Oxford, UK: Oxford University Press, 2003.
22. Verhelst ND, Glas CA, Verstralen HH. OPLM: Computer program and manual. Arnhem, The Netherlands: CITO, 1995.
23. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69.
24. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 2006;87:1388-95.

25. Ogince M, Hall T, Robinson K, Blackmore AM. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headache. Man Ther 2007;12:256-62.
26. Piva SR, Erhard RE, Childs JD, Browder DA. Inter-tester reliability of passive intervertebral and active movements of the cervical spine. Man Ther 2006;11:321-30.
27. Cleland JA, Childs JD, Fritz JM, Whitman JM, Eberhart SL. Development of a clinical prediction rule for guiding treatment of a subgroup of patients with neck pain: Use of thoracic spine manipulation, exercise, and patient education. Phys Ther 2007;87:9-23.
28. Tseng Y-L, Wang WTJ, Chen W-Y, Hou T-J, Chen T-C, Lieu F-K. Predictors for the immediate responders to cervical manipulation in patients with neck pain. Man Ther 2006;11:306-15.
29. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, Delitto A. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. Ann Intern Med 2004;141:920-8.
30. Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, Butler B, Garber M, Allison S. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27:2835-43.
31. Cramer G, Budgell B, Henderson C, Khalsa P, Pickar J. Basic science research related to chiropractic spinal adjusting: The state of the art and recommendations revisited. J Manipulative Physiol Ther 2006;29:726-61.
32. Legaspi O, Edmond SL. Does the evidence support the existence of lumbar spine coupled motion? A critical review of the literature. J Orthop Sports Phys Ther 2007;37:169-78.
33. Sizer PS, Brismée J-M, Cook C. Coupling behaviour of the thoracic spine: A systematic review of the literature. J Manipulative Physiol Ther 2007;30:390-9.
34. Brismée J-M, Atwood K, Fain M, Hodges J, Sperle A, Swaney M, Phelps V, Van Paridon D, Matthijs O, Sizer P. Interrater reliability of palpation of three-dimensional segmental motion of the lumbar spine. J Man Manip Ther 2005;13:215-20.
35. Panjabi M, Oxland T, Yamamoto I, Crisco J. Mechanical behaviour of the lumbar and lumbosacral spine as shown by three-dimensional load-displacement curves. J Bone Joint Surg Am 1994;76:413-24.
36. Harrison D, Harrison D, Troyanovich S. Three-dimensional spinal coupling mechanics : Part one. J Manipulative Physiol Ther 1998;21:101-13.

37. Maitland G, Hengeveld E, Banks K, English K (editors). Maitland's vertebral manipulation. 7th edition. Oxford, UK: Elsevier/Butterworth and Heinemann, 2005.
38. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.
39. Walker BF, Buchbinder R. Most commonly used methods of detecting spinal subluxation and the preferred term for its description: A survey of chiropractors in Victoria, Australia. J Manipulative Physiol Ther 1997;20:583-9.
40. Najm WI, Seffinger MA, Mishra SI, Dickerson VM, Adams A, Reinsch S, Murphy LS, Goodman AF. Content validity of manual spinal palpatory exams. A systematic review. BMC Complement Altern Med 2003;3:1.
41. Humphreys BK, Delahaye M, Peterson CK. An investigation into the validity of cervical spine motion palpation using subjects with congenital block vertebrae as a 'gold standard'. BMC Musculoskelet Disord 2004;5:19.
42. Abbott JH, McCane B, Herbison P, Moginie G, Chapple C, Hogarty T. Lumbar segmental instability: A criterion-related validity study of manual therapy assessment. BMC Musculoskelet Disord 2005;6:56.
43. Fernández-de-las-Peñas C, Downey C, Miangolarra-Page JC. Validity of the lateral gliding test as tool for the diagnosis of intervertebral joint dysfunction in the lower cervical spine. J Manipulative Physiol Ther 2005;28:610-6.
44. Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. Eur Spine J 2005;14:743-50.
45. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain. A systematic review of the literature. Spine 2004; 29:E413-25.
46. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: A systematic review. Aust J Physiother 2006;52:91-102.
47. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbæk L. Manual examination of the spine: A systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85.
48. Troyanovich SJ, Harrison DD. Motion palpation: It's time to accept the evidence. J Manipulative Physiol Ther 1998;21:568-71.
49. Haas M, Groupp E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay

assessment as an indicator for spinal manipulation. Spine 2003;28:1091-6.

50. Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: An examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehabil 2005;86:1745-52.
51. Kessler RC, Little RJA, Groves RM. Advances in strategies for minimizing and adjusting for survey nonrespons. Epidemiol Rev 1995;17:192-204.
52. Lusk C, Delclos GL, Burau K, Drawhorn DD, Aday LA. Mail versus internet surveys: Determinants of method of response preferences among health professionals. Eval Health Prof 2007;30:186-201.
53. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. J Clin Epidemiol 1997;50:1129-36.
54. Russell ML, Verhoef MJ, Injeyan HS, McMorland DG. Response rates for surveys of chiropractors. J Manipulative Physiol Ther 2004;27:43-8.
55. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089-92.

Appendix

Definitions of types of movements applied for passive intervertebral motion assessment of the spine

One-dimensional physiological movements: Moving one vertebra on another in the sagittal (flexion/extension), frontal (side-bending), or transverse (rotation) anatomical plane.

Three-dimensional physiological movements: Moving one vertebra on another in the sagittal, frontal and transverse anatomical planes simultaneously. An emphasis can be placed on any of these single components. Coupling of side-bending and rotation can be in the same direction or in opposite directions.

Accessory movements: Moving one vertebra on another using translatory motions associated with physiological motions.

Chapter 4b

The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual therapy: A qualitative interview study

Emiel van Trijffel, Thomas Plochg, Frank van Hartingsveld, Cees Lucas, Rob Oostendorp

J Man Manip Ther 2010;18:111-8

Abstract

Background: Passive intervertebral motion (PIVM) assessment is a characterising skill of manual therapists and is important for judgements about impairments in spinal joint function. It is unknown as to why and how manual therapists use this mobility testing of spinal motion segments within their clinical reasoning and decision-making.

Methods: This qualitative study aimed to explore and understand the role and position of PIVM assessment within the manual diagnostic process. Eight semistructured individual interviews with expert manual therapists and three subsequent group interviews using manual therapy consultation platforms were conducted. Line-by-line coding was performed on the transcribed data and final main themes were identified from subcategories. Three researchers were involved in the analysis process.

Results: Four themes emerged from the data: contextuality, consistency, impairment orientedness, and subjectivity. These themes were interrelated and linked to concepts of professionalism and clinical reasoning. Manual therapists used PIVM assessment within a multidimensional, biopsychosocial framework incorporating clinical data relating to the mechanical dysfunction as well as to personal factors while applying various clinical reasoning strategies. Interpretation of PIVM assessment and subsequent decisions on manipulative treatment were strongly rooted within practitioners' practical knowledge.

Conclusions: This study has identified the specific role and position of PIVM assessment as related to other clinical findings within clinical reasoning and decision-making in manual therapy in the Netherlands. We recommend future research in manual diagnostics to account for the multivariable character of physical examination of the spine.

Introduction

From early traditional international concepts in manual therapy, an emphasis has been placed on the diagnostics, treatment, and evaluation of joint function, especially of joints of the spine and pelvis.[1-4] A characterising feature of functional diagnostics is the use of passive joint movements of spinal motion segments for making judgements about the quality and quantity of segmental intervertebral joint function.[1] This passive intervertebral motion (PIVM) assessment is believed to play an important role within diagnostic clinical reasoning leading to classification of patients and treatment decisions.[5]

Systematic reviews have consistently shown low inter-examiner reliability for PIVM assessment.[6-11] In addition, the methodological quality of the studies reviewed was found to be poor and studies did not satisfy criteria for external validity disallowing generalisation of the results to clinical practice.[11] Most studies included non-representative participants, i.e., individuals who were not indicated to undergo PIVM assessment. Moreover, PIVM assessment has only been investigated as an independent factor within functional diagnostics which may not be reflective of clinical practice. However, it is unknown exactly what constitutes clinical practice in manual therapy with respect to the role of PIVM assessment within clinical decision-making in patients with spine-related disorders.

Recent surveys revealed that manual therapists (MT's) believe that findings from PIVM assessment, together with patient's history and other findings from physical examination, are important for deciding on manual therapy as a treatment option and that they are confident in their diagnostic conclusions drawn from PIVM assessment.[12;13] However, to date, an in-depth investigation into why and how MT's use PIVM assessment within their clinical reasoning has not been conducted.

This qualitative interview study was undertaken to explore why and how MT's use PIVM assessment within their clinical reasoning and decision-making. We aimed that its results could help guide the design and conduct of future studies into manual diagnostics leading to improved external validity of research results.

Methods

Study design

Data collection was based on individual and group interviews which have the advantage over paper-based cases of increasing the likelihood of revealing participants' reasoning as used in practice as opposed to their espoused theory.[14]

Objective and procedures

This qualitative study aimed to explore and understand the role and position of PIVM assessment within the manual diagnostic process. We appealed to the experiential knowledge of MT´s, expert teachers in manual therapy as well as clinicians, as a primary source of data collection. A purposive sample of 11 MT's was invited via email and a subsequent telephone call to participate in an individual interview. These therapists were all regarded as leading authorities within manual therapy covering the range of educational programs as acknowledged by the Royal Dutch Society for Physical Therapy (KNGF). Subsequently, nine groups of MT's constituting consultation platforms were invited to participate in group interviews. These platforms are part of the quality assurance program of the KNGF and generally consist of up to 15 therapists discussing quality improvement and assurance.[15] The majority of the platforms were established in 2002 and participation by therapists is geographically organised.

Participants

Three expert therapists (one with a Maitland background, one orthopedic manual therapist and one from the Master's program in Manual Therapy at the Vrije Universiteit Brussel (Free University of Brussels, Belgium) declined to participate in the individual interviews because of time constraints. Characteristics of the remaining eight participating experts are summarised in Table 1. The majority of participants were highly experienced in practising as well as in teaching manual therapy.

Four platforms agreed to participate in group interviews of which three were initially used for data collection. Of the remaining five platforms, three could not participate due to lack of time and two did not respond to our invitation. Characteristics of the three participating groups are presented in Table 2.

Table 1. Demographic and professional characteristics of expert manual therapists participating in the individual interviews *(n = 8)*

Participant	Gender	Age (yr)	Experience in MT (yr)	Experience in MT teaching (yr)	MT background
1	m	42	14	0	SOMT
2	m	58	30	30	SOMT
3	m	51	22	21	SOMT
4	m	47	16	16	SOMT
5	m	52	18	18	VUB
6	m	33	8	8	VUB
7	f	49	22	15	Maitland
8	m	56	29	27	OMT

f: female, m: male, MT: manual therapy, OMT: Orthopaedic Manual Therapy, SOMT: Stichting Opleidingen Musculoskeletale Therapie (Institute for Master Education in Musculoskeletal Therapy), VUB: Vrije Universiteit Brussel (Free University of Brussels), Master Manual Therapy, Belgium

Data collection

Individual interviews with the eight experts took place between November, 2007 and April, 2008. Interviews were conducted by the principal researcher (EvT) who is an experienced manual therapist and trained as a qualitative researcher. Interviews were semistructured and an interview guide was used that contained the following topics exploring key aspects of clinical reasoning within manual diagnostics: (1) the use of PIVM assessment as related to findings from patient’s history and other clinical tests; (2) the interpretation of clinical findings from PIVM assessment; (3) the role of PIVM assessment in selecting manual therapy as a treatment option; (4) required knowledge and skills for using and interpreting PIVM assessment; (5) the role of PIVM assessment within a biopsychosocial approach; and (6) the importance of PIVM assessment for the identity of manual therapy. Interviews were audio-recorded and the interviewer made additional notes of specific quotes and observations. Interview time ranged from 50 to 75 minutes. The purpose of these interviews was to cover a wide range of perspectives on the role and position of PIVM assessment within clinical reasoning and decision-making across various manual therapy approaches. It was decided in

Table 2. Demographic and professional characteristics of manual therapy consultation platforms participating in the group interviews *(n = 3)*

Group	Number of participants	Gender (males)	Age* (yr)	Experience in MT practice* (yr)	MT background
1	8	7	37.5 (31-49)	5.5 (3-13)	SOMT (n = 8)
2	11	6	48 (37-63)	12 (5-23)	SOMT (n = 5), OMT (n = 5), MT Utrecht (n = 1)
3	8	7	45 (40-55)	13 (8-16)	SOMT (n = 8)

MT: manual therapy, OMT: Orthopaedic Manual Therapy, SOMT: Stichting Opleidingen Musculoskeletale Therapie (Institute for Master Education in Musculoskeletal Therapy)
* presented as median (minimum-maximum)

advance that a fixed number of interviews would suffice. Between interviews, the interviewer repetitively reflected on his role as an interviewing manual therapist in order to reduce researcher bias. In addition, he was peer reviewed by a second researcher (FvH), who specifically addressed issues such as leading questions and interviewer's prejudice. Member checking was performed to enhance the validity of the raw transcribed material first and, subsequently, of analysed data as well.

Group interviews took place between June, 2008 and September, 2008. EvT conducted the interviews using a topic list similar to the one used in the individual interviews. Elicitation exercises are helpful in focusing the groups' attention on the study topic and allow comparative analysis.[16] A ranking exercise was used to facilitate participants' thinking about using PIVM assessment within their reasoning in a case of non-specific mechanical neck pain in which few demographic (age, gender) and clinical (duration of complaints, localisation of pain) data were given. In this exercise, participants were requested to reach consensus about the order in which they would apply clinical examination tests with specific attention to the role of PIVM assessment. The therapists were encouraged to share how they would think and act in this case in daily practice instead of how they should think and act. Interviews were audio-recorded and the interviewer made additional notes of specific quotes, observations, and interaction between participants. Each interview lasted 90 minutes. The purpose of these interviews was to test whether themes and categories from analysed individual interviews could be identified in groups of therapists representing clinical practice in manual therapy. Saturation of data was used to determine the number of interviews required. FvH peer reviewed the interview process.

Informed consent was obtained from all participants and their anonymity was ensured by allocating numbers instead of using their names during analysis. In addition, confidentiality of data was ensured.

Data analysis

All taped interview data were transcribed verbatim. Analysis took place after every interview. Line-by-line open coding was performed by the principal investigator and identified codes were classified into categories. Two researchers (EvT, FvH) discussed the labeling of categories until agreement was reached. During the process of labeling and analysis, both researchers independently explored the data in search of deviant cases and disconfirming data. Through discussion and consensus, emerging final main themes were agreed upon by three researchers (EvT, TP, and FvH). Subsequently, themes were further integrated by incorporating a sociological theory of professionalism[17] as well as a biopsychosocial, collaborative hypothesis-oriented model of clinical reasoning as described by Jones et al.[18] Quotes were selected illustrating each category and translated with the help of a native speaker. Throughout the research process, EvT kept a logbook and made memos to record changes in methods and decisions regarding data collection and analysis.

Results

From the analysis of the individual interview data, four themes emerged: contextuality, consistency, impairment orientedness, and subjectivity. These themes were, to a large extent, corroborated by findings from the group interviews. The Figure illustrates how the four themes are interrelated and are linked to various types of clinical reasoning strategies. Professionalism acts as a covering main theme.

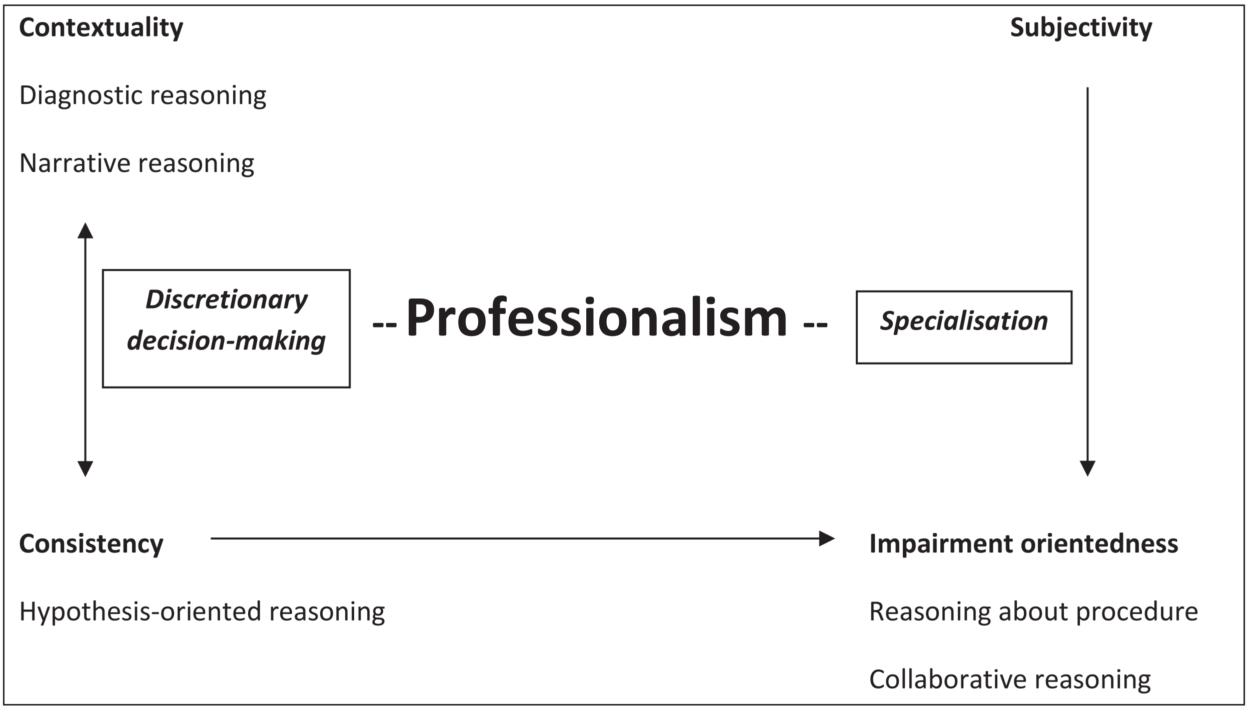

Figure. Illustration of the relationship between the four themes (contextuality, consistency, subjectivity, impairment orientedness) with each other, containing elements of certain strategies for clinical reasoning (as described by Jones et al[18]), as well with two key elements of professionalism: discretionary decision-making and specialisation[17]

Below, a more detailed description of the results is given for the individual and group interviews separately and themes are illustrated by quotes.

Individual interviews

Throughout the interviews, expert MT's demonstrated a high level of concern by enthusiastically expressing their firm visions on manual therapy profession and education. Afterwards, member checking rounds did not generate additional comments.

Contextuality

Respondents argued that the indication for using PIVM assessment is dictated by findings from patient's history as well as from other clinical tests. They believe that the patient's personal perspectives and characteristics are important for deciding on PIVM assessment besides information about movement-related impairments and activity limitations. Within this multidimensional context, the patient's history is a decisive source of information that guides further collection of clinical data and, more specifically, the use of PIVM assessment, which is illustrated by a statement from Respondent 2 (R2):

> *So, in general, to identify signs from patient's history which would indicate the use of passive segmental motion examination, that patient HAS to have told me 'I have restricted activities, like looking over my shoulder or bending forward,' such that make me consider the existence of impairments in mobility.* (**R2**)

In addition, other motion examination findings are considered before using PIVM assessment; however, PIVM assessment seems to be used routinely.

> *To me, when it is a non-specific problem, and it is a mechanical one, I will definitely use it [PIVM assessment] [........] ALWAYS.* (**R4**)

From the previous, it may appear that deciding on PIVM assessment, although depending on findings from patient's history and other clinical tests, is predominantly led by mechanical arguments. However, all eight experts did reason about an indication for using PIVM assessment from other perspectives as well. In particular, they explicitly include personal factors related to the patient's behaviour and beliefs in their decision-making thereby adopting a biopsychosocial approach to manual diagnostics. Among other factors mentioned were duration of complaints, pain intensity, muscular defense, physical fitness and fatigue, posture and working positions, and accompanying neurogenic complaints.

> *When I'm suspicious, after taking the patient's history, of other aspects contributing to movement dysfunction, like in the case of chronic benign pain, then there is NO reason to perform passive segmental motion examination.* (**R3**)

Consistency

Interviewees stated that they use PIVM assessment during manual examination in order to check and confirm earlier clinical findings. Implicitly, they generate hypotheses about correlations between what they were told by patients and what they found during physical examination. PIVM assessment, then, plays a role in confirming the presence or absence of impairments in spinal joint motion that can be related to the patient's pain, activity limitations, and participation restrictions. Respondent 1, however, was reticent in giving credence to the significance of the findings from PIVM assessment, because he takes into account the lack of scientific evidence for PIVM assessment. This issue of the importance of available evidence for PIVM assessment was subsequently added as a topic during the remaining interviews. Experts had differing opinions in this respect. Some (R3 and R4) applied a very pragmatic approach. For example,

> *I am aware of the lack of evidence. It just isn't there but that doesn't influence my daily practice [........] I am convinced that whatever we do we should continu like we are [........] waiting for evidence just takes too long. It's a shame that inevitably sometimes you do things that are not helpful [........] so be it.* (**R4**)

By tailoring diagnostics to individual patients, therapists employ a high level of autonomy in their reasoning and decision-making. This discretionary decision-making is believed to be a crucial element of a manual profession.[17] Data supporting the two themes of contextuality and consistency imply a certain order in conducting tests during manual examination. Indeed, all respondents admitted to a more or less fixed order in which PIVM assessment comes in later or even last. It was decided to explore this issue further as a main focus in the group interviews.

Subjectivity

Subjectivity refers to the lack of objective measures for interpreting and classifying clinical findings from PIVM assessment. Variation in interpretation of quality and quantity of intervertebral motion is an inevitable consequence of therapists' own clinical experience from which their individual frame of reference is built.

Manual therapy is a craft really that you have to learn and that is built up through experience, I think. I can read about it but learning to interpret test findings I think you have to learn on the job. (**R6**)

One respondent (R8), however, stated that he uses PIVM assessment as an objective measure by comparing its findings with 'real' subjective ones, namely, those reported by patients themselves, and he believes this is actually a strong feature of manual therapy.

The experts recognised that lack of uniformity in criteria for judging impairments of spinal motion segments hinders the profession's transparency towards patients and referrers, and they explicitly recommend, most of them being teachers, thorough training of students by experienced practitioners in order to reach more consensus on how to judge and express impairments of the functions of spinal motion segments.

Impairment orientedness

The presence of impairments in spinal joint function among consistent clinical findings guided the decision of selecting manual therapy, either mobilisation or manipulation, as a treatment option. The experts fully agreed that the skills for diagnosing and treating spinal joint motion impairments are a distinct feature of manual therapy and as such separate the manual therapy competency domain from that of physiotherapy. Manual therapy has a strong focus on knowledge of joint arthrokinematics and osteokinematics and on impairments of joint function and, as treatment is aimed at individual spinal segmental levels, PIVM assessment is necessary for decisions about which motion segment to treat and how to treat. Respondents 1 and 5, however, took a critical view, reflecting on the limitations of this narrow focus for the profession:

R5: *I believe manual therapy suffers from an inflated ego*.

Interviewer: *What do you mean?*

R5: *The simplifying of the patient's complaints into segmental dysfunctions and the assumption that removing these dysfunctions will automatically lead to the patient's recovery.*

It was striking how even expert teachers in manual therapy were not able to put into words how and which clinical findings from PIVM assessment would lead to a choice for either mobilisation or manipulation of a joint. Type of end-feel, amount of restricted motion, number of motion segments involved, level of patient's pain intensity, but also characteristics of the patient and his or her former experience with manual therapy are factors considered in deciding on a manipulative intervention. In conclusion, the choice for the type of intervention seems to be multidimensionally determined and influenced by therapists' own subjective preferences and experience as part of their individual practical knowledge.

Group interviews

In the given case of non-specific mechanical neck pain, all three groups of therapists reached consensus on the sequence of testing procedures for manual examination. Moreover, there was complete agreement on this ordering between groups. After history taking and inspection, active motion assessment, passive regional motion assessment, and passive segmental motion assessment are applied respectively, which, depending on findings and not always during the same first session, could be followed by muscle function examination and neurodynamic evaluation. The groups also indicated that the decision for applying PIVM assessment depends on earlier clinical findings, either related to the mechanical problem or to patient's external or personal factors. Although participants admitted to using PIVM assessment for checking and confirming patient's complaints, they had difficulty explaining how this relates to the position of this assessment following other tests. The following fragment, containing a discussion between four participants (P) in Group 3, illustrates how strongly education prescribes acting by professionals in practice:

> **Interviewer**: *Why is passive intervertebral motion assessment positioned last in line?*
>
> **P3**: *That's what we are used to doing.*
>
> **P6**: *In a pyramid in which you start broadly with history taking, you enter some sort of funnel model and you go on getting more specific, and segmental motion assessment is as specific as you can get.*
>
> **P1**: *It is an automatic activity of steps you pass through as a rule.*
>
> **P7**: *Yeah, I believe that's what we've been taught.*
>
> **Interviewer**: *How come?*

P3: *That originates from the structure that is handed to you during training.*

Participants in all groups could not agree on whether PIVM assessment should be judged primarily on function (i.e., mobility or stability) or on pain provocation and, even more challenging, when judged on both, which judgement should come first during testing. It was notable that participants in Group 1, being younger and more recently trained, perceive their reasoning skills as more important than their physical examination skills when asked about the additional value of manual therapy as compared to physiotherapy. On the other hand, the more experienced therapists in Groups 2 and 3 expressed a more patient-centered approach by consciously using findings from PIVM assessment for educating patients and involving these findings in choosing and evaluating patient management.

Given the similarities of opinions and disagreements across the three groups of practitioners, we decided that the remaining fourth available consultation platform would not be used for further exploration.

Discussion

This qualitative study has been the first to shed light on the mental processes of clinical reasoning and decision-making by MT's as related to PIVM assessment and has provided level 5 evidence for the role and position of this test procedure within the manual diagnostic process.[19] Identifying the role and position of a test within a diagnostic strategy helps design studies to evaluate the diagnostic value of tests.[20] In diagnostic research, a stepwise evaluation of tests is increasingly proposed to consider not only the test's technical accuracy but also its place in the clinical pathway and, eventually, its impact on patient outcomes.[21]

We found that PIVM assessment is positioned, albeit sometimes more or less routinely, as an 'add-on' test after history taking, visual inspection, and active and passive motion examination. Add-on tests are generally used to increase the sensitivity or specificity of a diagnostic strategy in order to improve treatment selection.[20;22] Increased sensitivity through adding PIVM assessment could identify patients with segmental joint hypomobility newly indicated for, say, manipulative treatment in the absence of active motion restrictions or activity limitations. Increased specificity limits the number of false-positive diagnostic conclusions and would confirm an indication for treatment in those patients already testing positive on preceding motion examination and activity limitations. Research results are in favour of the latter, demonstrating higher levels of specificity for spinal motion segment testing as compared to its sensitivity.[23-27] However, to date, research on PIVM assessment can be regarded as test research following a single-test or univariable approach thus neglecting the multivariable character of diagnostics as opposed to diagnostic research.[28]

Our data support a multivariable, biopsychosocial approach to research into manual diagnostics in general and PIVM assessment in particular. De Hertogh et al[29] showed improved accuracy of manual examination of cervical motion segments when clustered with results on pain intensity and medical history, and claimed that this multidimensional approach better resembles practice. The reliability and, if possible, accuracy of either add-on diagnostic strategy as a whole should be the focus of future research including representative patients who are indicated to undergo PIVM assessment and potentially yielding study results more reflective of diagnostic pathways used in clinical practice. A proposed research objective could be to determine the inter-examiner reliability of intervertebral mobility testing of impaired motion segments, identified through reliable pain provocation tests[9], in patients with either spine-related complaints or extremity disorders indicated to undergo spinal examination after testing negative on 'yellow flags' but showing active range of motion restrictions and activity limitations during history taking and physical examination. At some point, studies inevitably need to incorporate patient outcomes while evaluating test-plus-treatment strategies.[22]

Previous research investigating clinical reasoning in the domain of musculoskeletal physiotherapy focused on exploring characteristics of expert practitioners and indicated the use of various diagnostic reasoning processes such as pattern recognition, hypothetico-deductive reasoning, and patient-centered, collaborative reasoning.[30-38] MT's indeed apply a hypothetico-deductive approach in their encounters with patients.[38] These results seem in contrast with findings from research in doctors showing a pattern recognition mode of reasoning as clinical expertise grows.[39] However, it is now recognised that clinicians, often unconsciously, use multiple combined strategies of reasoning to solve clinical problems.[40] Already in undergraduate students, conceptualizations of clinical reasoning in musculoskeletal physiotherapy ranged from relatively simple to increasingly complex but mixed forms of reasoning.[41]

Our respondents, expert teachers as well as practising clinicians, could not agree on which clinical finding is indicative for dysfunctions of spinal motion segments or directive for decisions on manual treatment. Maher et al[42] showed that MT's conceptualise spinal stiffness in an individual, multidimensional manner, and joint and tissue characteristics are described in qualitative terms. The highly subjective interpretation of PIVM assessment is embedded within and contributes to the practical craft knowledge characterising the profession.[43] However, it may also account for its low reliability.[42] De Hertogh et al[29] chose a more pragmatic approach by marking manual examination as positive when at least any two out of three criteria (mobility, end-feel, pain provocation) were met. They showed improved reliability and high specificity of manual examination in neck pain patients confirming earlier findings by Jull et al.[5,29,44] Combined interpretation of findings from PIVM assessment, clustered with other signs and symptoms, looks to be a promising approach to future research on the reliability and diagnostic accuracy of manual diagnostics leading to transferable results.

Dutch MT's believe that PIVM assessment is important for deciding on a treatment strategy.[13] Authors have questioned the clinical usefulness and necessity of identifying impairments of joint mobility at specified spinal levels in order to make treatment decisions.[45-48] Seffinger et al[9] concluded that assessing regional range of spinal motion is more reliable than segmental examination. On the other hand, Chiradejnant et al[49] showed a greater reduction in pain intensity when mobilisation was applied to the symptomatic lumbar motion segment rather than to a randomly assigned level. Despite the limited evidence for a spinal motion segment approach, Dutch MT's derive their status as specialists in the care of spine-related health problems, as opposed to non-specialised physiotherapists, in great part from their skill to address manual diagnostics and treatment to individual spinal motion segments.

Finally, the large amount of agreement between and among our respondents was remarkable. Despite the fact that therapists trained in the largest manual therapy educational institute in the Netherlands (SOMT: Stichting Opleidingen Musculoskeletale Therapie) were overrepresented in both samples, our expert teachers still had different educational backgrounds representing different manual therapy approaches. It may be concluded that the various concepts of manual therapy still share many common sources of knowledge dating back to the early origins of the profession.[50] From their Delphi study among US manual therapy educators, Sizer et al[51] identified consensual skill sets associated with competent application of orthopedic manual therapy despite the disparate backgrounds of respondents. Manual joint assessment was contained in the majority of stand-alone descriptor statements.[51] In addition, Maher et al[42] found similar results between US and Australian manipulative physiotherapists for the conceptualisation of spinal stiffness.

Limitations of this study

Although interviews are the most common method for producing qualitative data, a shortcoming is that they provide access to what people say they think and do, not what they actually think and do.[52] Furthermore, the principal expert investigator was the conductor of all interviews. Collected data could have been shaped by the influence of his prior assumptions and experience, and these could have introduced personal and intellectual biases into the results. However, we believe that using an explicit topic list during the interviews and taking a reflexive position towards data collection and analysis, including peer review, have sufficiently protected against biased interpretation of results by the conductor.

With respect to the external validity of our results, we point to the specific system for manual therapy education in the Netherlands, where manual therapy is considered a post-graduate (non-university) specialisation following entry-level bachelor physiotherapy education and education programs that meet the Educational Standards of the International Federation of Orthopaedic Manipulative Therapists.[53] We fully acknowledge that the Dutch educational framework may strongly differ from that in other countries, like the USA, Canada, and Australia, in which specific knowledge and manual skills for diagnosing and treating spinal segmental joint impairments is entry-level. Therefore, our results based on the verbal expressions of our respondents may not always apply beyond the Dutch population of MT's.

Finally, we included a purposive sample of expert MT's to cover the range of different perspectives on the study subject from the various manual therapy educational programs acknowledged for registration in the Netherlands. However, we did not aim for data saturation in this part of the study and, therefore, we could not search for deviant cases and contradicting opinions further within every single approach.

Conclusions and recommendations

This study has provided insight into why and how PIVM assessment is used by Dutch MT's within their clinical reasoning and decision-making. In addition, the specific role and position of mobility testing of spinal motion segments, as related to the patient's history and other clinical tests, has been exposed. We recommend future research into manual diagnostics to account for the multivariable, biopsychosocial, hypothesis-oriented character of physical examination of the spine and of PIVM assessment in particular.

References

1. Farrell JP, Jensen GM. Manual therapy: A critical assessment of role in the profession of physical therapy. Phys Ther 1992;72:843-52.
2. Maher C, Latimer J. Pain or resistance – the manual therapists' dilemma. Aust J Physiother 1992;38:257-60.
3. NVMT (Dutch Association for Manual Therapy). Professional competence profile for the manual therapist (May 2005). Amersfoort, The Netherlands. Available at: http://www.nvmt.nl/upload/BCP.ENG.pdf. (accessed 31 January, 2009)
4. Van Ravensberg CDD, Oostendorp RAB, Van Berkel LM, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual therapy: Differences in patient characteristics. J Man Manip Ther 2005;13:113-24.
5. Jull G, Treleaven J, Versace G. Manual examination: Is pain provocation a major cue for spinal dysfunction? Aust J Physiother 1994;40:159-65.
6. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
7. Hestbæk L, Leboeuf-Yde C. Are chiropractic tests for the lumbo-pelvic spine reliable and valid? A systematic critical literature review. J Manipul Physiol Ther 2000;23:258-75.
8. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain. Aust J Physiother 2006;52:91-102.
9. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain. A systematic review of the literature. Spine 2004; 29:E413-25.
10. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbæk L. Manual examination of the spine: A systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85.
11. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69.
12. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.

13. Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine. A survey of Dutch physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51.
14. Argyis C, Schön D. Theory in practice. Increasing professional effectiveness. San Francisco, CA: Jossey-Bass, 1974.
15. Van der Wees PJ, Hendriks EJM, Veldhuizen RJ. Quality assurance in the Netherlands: From development to implementation and evaluation. Dutch Journal of Physical Therapy (Special Issue International WCPT Congress Barcelona 2003) 2003;113(S):3-6.
16. Colluci E. "Focus groups can be fun": The use of activity-oriented questions in focus group discussions. Qual Health Res 2007;17:1422-33.
17. Freidson E. Professionalism. The third logic. Oxford, UK: Blackwell Publishers Ltd, 2001.
18. Jones MA, Jensen G, Edwards I. Clinical reasoning in physiotherapy. In: Higgs J, Jones MA, Loftus S, Christensen N (editors). Clinical reasoning in the health professions (3rd edition). Edinburgh, UK: Elsevier/Butterworth Heinemann, 2008.
19. CEBM (Oxford Centre for Evidence Based Medicine). Levels of evidence (March 2009). Available at: http://www.cebm.net/index.aspx?o=1025. (accessed 17 May, 2009)
20. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089-92.
21. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F. The evaluation of diagnostic tests: Evidence on technical and diagnostic accuracy, impact on patient outcome and cost effectiveness is needed. J Clin Epidemiol 2007;60:1116-22.
22. Lord SJ, Irwig L, Bossuyt PMM. Using the principles of randomized controlled trial design to guide test evaluation. Med Decis Making 2009;29:E1-12.
23. Abbott JH, McCane B, Herbison P, Moginie G, Chapple C, Hogarty T. Lumbar segmental instability: A criterion-related validity study of manual therapy assessment. BMC Musculoskelet Disord 2005;6:56.
24. Fritz JM, Whitman JM, Childs JD. Lumbar spine segmental mobility assessment: An examination of validity for determining intervention strategies in patients with low back pain. Arch Phys Med Rehab 2005;86:1745-52.
25. Hall TM, Robinson KW, Fujinawa O, Akasaka K, Pyne EA. Intertester reliability and diagnostic validity of the cervical flexion-rotation test. J Manipulative Physiol Ther 2008;31:293-330.

26. Humphreys BK, Delahaye M, Peterson CK. An investigation into the validity of cervical spine motion palpation using subjects with congenital block vertebrae as a 'gold standard'. BMC Musculoskelet Disord 2004;5:19.
27. Ogince M, Hall T, Robinson K, Blackmore AM. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headache. Man Ther 2007;12:256-62.
28. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem 2004;50:473-6.
29. De Hertogh WJ. Vaes PH, Vijverman V, De Cordt A, Duquet W. The clinical examination of neck pain patients: The validity of a group of tests. Man Ther 2007;12:50-5.
30. Doody C, McAteer M. Clinical reasoning of expert and novice physiotherapists in an outpatient orthopaedic setting. Physiotherapy 2002;88:258-68.
31. Edwards I, Jones M, Carr J, Braunack-Mayer A, Jensen GM. Clinical reasoning strategies in physical therapy. Phys Ther 2004;84:312-35.
32. Jensen GM, Shepard KF, Hack LM, Gwyer J. Attribute dimensions that distinguish master and novice physical therapy clinicians in orthopaedic settings. Phys Ther 1992;72:711-22.
33. Jensen GM, Gwyer J, Shepard KF, Hack LM. Expert practice in physical therapy. Phys Ther 2000;80:28-43.
34. King CA, Bithell C. Expertise in diagnostic reasoning: A comparative study. Int J Ther Rehabil 1998;5:78-87.
35. May S, Greasley A, Reeve S, Withers S. Expert therapists use specific clinical reasoning processes in the assessment and management of patients with shoulder pain: A qualitative study. Aust J Physiother 2008;54:261-6.
36. Payton OD. Clinical reasoning process in physical therapy. Phys Ther 1985;65:924-8.
37. Resnik L, Jensen GM. Using clinical outcomes to explore the theory of expert practice physical therapy. Phys Ther 2003;83:1090-1106.
38. Rivett DA, Higgs J. Hypothesis generation in the clinical reasoning behaviour of manual therapists. J Phys Ther Educ 1997;11:40-45.
39. Kempainen RR, Migeon MB, Wolf FM. Understanding our mistakes: A primer on errors in clinical reasoning. Acad Med 2003;25:177-81.
40. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med 2006;355:2217-25.

41. Hendrick P, Bond C, Duncan E, Hale L. Clinical reasoning in musculoskeletal practice: Students' conceptualizations. Phys Ther 2009;89:430-42.
42. Maher CG, Simmonds M, Adams R. Therapists' conceptualization and characterization of the clinical concept of spinal stiffness. Phys Ther 1998;78:289-300.
43. Titchen A, Ersser SJ. The nature of professional craft knowledge. In: Higgs J, Titchen A (editors). Practice knowledge and expertise in the health professions. Oxford, UK: Butterworth Heinemann, 2001.
44. Jull G, Zito G, Trott P, Potter H, Shirley D, Richardson C. Inter-examiner reliability to detect painful upper cervical joint dysfunction. Aust J Physiother 1997;43:125-9.
45. Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. Eur Spine J 2005;14:743-50.
46. Haas M, Groupp E, Panzer D, Partna L, Lumsden S, Aickin M. Efficacy of cervical endplay assessment as an indicator for spinal manipulation. Spine 2003;28:1091-6.
47. Huijbregts PA. Spinal motion palpation: A review of reliability studies. J Man Manip Ther 2002;10:24-39.
48. Troyanovich SJ, Harrison DD. Motion palpation: It's time to accept the evidence. J Manipulative Physiol Ther 1998;21:568-71.
49. Chiradejnant A, Latimer J, Maher CG, Stepkovitch N. Does the choice of spinal level treated during posteroanterior (PA) mobilisation affect treatment outcome? Physiother Theory Pract 2002;18:165-74.
50. Pettman E. A history of manipulative therapy. J Man Manip Ther 2007;15:165-74.
51. Sizer PS, Felstehausen V, Sawyer S, Dornier L, Matthews P, Cook C. Eight critical skill sets required for manual therapy competency: A Delphi study and factor analysis of physical therapy educators of manual therapy. J Allied Health 2007;36:30-40.
52. Green J, Thorogood N. Qualitative methods for health research. London, UK: Sage Publications, 2004.
53. IFOMT (International Federation of Orthopaedic Manipulative Therapists). Educational standards in orthopaedic manipulative physical therapy. Part A: Educational standards (11 June 2008). Available at: http://www.ifomt.org/pdf/IFOMT_Education_Standards_and_International_Monitoring_20080611. pdf. (accessed 31 January, 2009)

Chapter 5

Manual therapists' use of biopsychosocial history taking in the management of patients with back or neck pain in clinical practice

Rob Oostendorp, Hans Elvers, Emilia Mikolajewska, Marjan Laekeman, Emiel van Trijffel, Han Samwel, William Duquet

ScientificWorldJournal 2015;2015:170463

Abstract

Background: Little is known about the extent to which manual therapists apply the biopsychosocial concept in their process of clinical reasoning in patients with musculoskeletal pain, in particular non-specific back or neck pain.

Methods: The SCEBS method (Dutch: SCEGS methode), covering the Somatic, psychological (Cognition, Emotion, Behaviour), and Social dimensions of chronic pain, was used to evaluate biopsychosocial history taking by manual therapists. In Phase 1, process indicators were developed while in Phase 2 these indicators were tested in practice.

Results: Literature-based recommendations were transformed into 51 process indicators. Twenty manual therapists contributed 108 patient audio recordings. History taking was excellent (98.3%) for the Somatic dimension, very inadequate for Cognition (43.1%) and Behaviour (38.3%), weak (27.8%) for Emotion, and low (18.2%) for the Social dimension. Manual therapists estimated their coverage of the Somatic dimension as excellent (100%), as adequate for Cognition, Emotion, and Behaviour (60.1%), and as very inadequate for the Social dimension (39.8%).

Conclusions: Manual therapists perform screening for musculoskeletal pain mainly through the use of the somatic dimension of (chronic) pain. The psychological and social dimensions of chronic pain were inadequately covered. Furthermore, a substantial discrepancy between actual and self-estimated use of biopsychosocial history taking was noted. We strongly recommend implementation of the SCEBS method in educational programs in manual therapy.

Introduction

Since the introduction of the biopsychosocial disease model by Engel[1], there has been a considerable shift in the use of this model for the diagnosis and management of musculoskeletal disorders such as back and neck pain. In the past, the biomedical model predominantly focused on anatomical structures related to the back and neck region as the origin of pain and as justification for medical interventions. The subsequent failure of many treatment approaches, amongst other factors, highlighted the limitations of the biomedical model in the treatment of patients with musculoskeletal disorders.

Together with contributions from many other similar papers, a publication by Waddell[2] in 1987 in particular catalysed the worldwide introduction of the biopsychosocial model for patients with spinal disorders. The last 40 years have seen a surge in research on neuro- and behavioural sciences including those related to the field of manual therapy.[3-5] This has led to a greater appreciation of the role of psychological and social factors that impact (chronic) musculoskeletal pain. A number of factors, including the high incidence and prevalence of patients with chronic musculoskeletal pain, the accumulating evidence supporting a role for psychological and social factors in relation to chronic pain, the increasing number of clinical practice guidelines based on scientific evidence, the international classifications (e.g., International Classification of Functioning, Disability and Health (ICF)[6], and the growing interest in the clinical reasoning process, point to the relevance of a broader approach to the management of patients with musculoskeletal disorders. Despite this, little is still known about the extent to which manual therapists apply the biopsychosocial concept in their process of clinical reasoning for patients with musculoskeletal pain, particularly non-specific back and neck pain.

The process of clinical reasoning consists of a diagnostic phase (history taking, [objectives of] physical examination, analysis, and conclusion), a treatment phase (treatment plan and treatment), and an evaluation phase (evaluation and discharge). The 'history taking' is the first step in the diagnostic phase and is crucial to the orientation on the health problem of patients with (chronic) musculoskeletal pain in terms of (impairments in) body functions and structures, activity (limitations), participation (restrictions), and personal and environmental factors. The SCEBS method (Dutch: SCEGS methode), developed in 1995 by medical psychologists Van Spaendonck and Bleijenberg, was designed as a diagnostic framework for general practitioners who are less familiar with the biopsychosocial history taking in patients with (chronic) pain.[7-9] This method identifies three dimensions of pain: the Somatic or biological dimension, the psychological dimension (Cognition, Emotion and Behaviour), and the Social dimension. A set of sample questions was developed for each dimension such as ''Can you move your back/neck?" (to trace impairments of movement-related functions), "What do you think when you are experiencing pain?" (to trace catastrophic or helplessness cognitions, fear of pain, lack of self-efficacy, or unrealistic treatment expectations), "How do you feel when you experience pain?" (to trace depression or anxiety), "What do you do in response of pain?" (to trace avoidance behaviour or pain resistance behaviour), and "How does your social environment react to your pain?" (to trace maladaptive social responses to pain behaviour). The SCEBS method is commonly used in the Netherlands by general practitioners, occupational physicians, psychologists, nurses, and, to a lesser extent, manual therapists, for the initial orientation and analysis in patients with inexplicable and unexplained pain.[10;11]

The transparency of the SCEBS method-based process of history taking using measurable elements such as quality indicators (QIs) is seen as one of the cornerstones of the quality of care, particularly the quality of manual therapy in patients with (chronic) musculoskeletal pain. QIs have been defined as 'measurable elements of practice performance for which there is evidence or consensus that they can be used to assess the quality, and thus change the quality, of care provided'.[12;13] QIs are related to structures (such as staff, equipment, and appointment systems), processes (such as prescribing, investigations, and clinical reasoning), and outcomes (such as mortality, morbidity, patient satisfaction, and functioning) of care.[14] QIs are preferably derived from guideline-based recommendations supplemented by expert clinical experience and patient perspectives, and developed by means of a systematic method.[15] After development, sets of QIs should be subjected to a pilot practice test.

The present study focused on the development and evaluation of process indicators in patients with chronic musculoskeletal pain, with an emphasis on non-specific back and neck pain. The two primary goals of this study were (1) to develop a set of process indicators relevant to biopsychosocial history taking in patients with non-specific back and neck pain and (2) to subject this set to a pilot practice test to determine its value in assessing the actual extent of implementation of biopsychosocial history taking in Dutch manual therapy practice.

Methods

Design

The study consisted of two phases: (i) indicator development and (ii) indicator testing through a pilot practice test. The QI development included three steps: (i) extraction of recommendations from the original description of the SCEBS method and relating literature, (ii) transformation of recommendations into process indicators, and (iii) classification of process indicators according to the SCEBS method. For the pilot practice test, we used a cross-sectional design to test the integration of biopsychosocial history taking in manual therapy practice.

The Medical Ethics Committee of the Radboud University Medical Centre, Nijmegen, the Netherlands, stated in writing that ethical approval was not necessary. Each practice formally consented to participate and all patients were informed about the study and gave permission for anonymous use of data.

Phase 1: Indicator development

Step 1: Extraction of recommendations

Recommendations were identified using the original SCEBS method literature, systematic reviews of the screening, assessment, and management of patients with non-specific back or neck pain, and ICF core sets for musculoskeletal disorders. These recommendations were extracted by two members of the research team (RO and WD) and, where necessary, differences were discussed with a third member until consensus was reached. Based on these recommendations, a set of questions was formulated for each dimension (e.g., the Somatic dimension: what are the type, localisation, intensity, frequency, and duration of pain, and what are the impairments of neuromusculoskeletal and movement-related functions (such as mobility and stability of joint functions); Cognition: what are your expectations of treatment?).

Step 2: Transformation into process indicators

The questions were transformed into process indicators by treating them as percentages of patients who were asked a certain question (i.e., the percentage of patients who were asked specific questions about their attributions of pain).

Step 3: Classification of process indicators

The process indicators were classified into the three dimensions of the SCEBS method with eight subdimensions of the psychological dimension (Appendix).

Phase 2: Indicator testing

An invitation to participate in the pilot practice test was sent to 112 physiotherapy practices in the south of the Netherlands of which 68 (60.7%) indicated interest (Figure).

From the 68 practices, manual therapists from 49 (72.1%) practices participated in a regional information session that outlined the purpose and content of the study and the expected contribution. Of the 49 practices, 27 (55.1%) enrolled 21 manual therapists. These manual therapists were asked to collect data on at least five new patients with non-specific back or neck pain, preferably on the first new patient each week meeting the criteria. Based on the number of participating manual therapists, the number of patients expected to be included in the study was about 100. Patients had to meet the following inclusion criteria: aged 18 to 70 years, pain and/or stiffness in the back or neck for at least six weeks, back or neck complaints reproducible during active or passive examination, and written informed consent. Non-specific back or neck pain was defined as pain with no specific cause such as systemic disease, fracture, or other organic disorders. Patients with a history of additional complaints such as non-radicular pain were included only if the back or neck pain was dominant. Patients whose history, signs, and symptoms suggested a potential non-benign cause (including previous surgery of the back or neck) or those who showed evidence of a specific condition such as malignancy, neurologic disease, fracture, herniated disc, or systemic rheumatic disease were excluded.

Data collection

Data were collected over a period of six months. The history taking during the first appointment took place in the manual therapists' practice and was recorded using digital audio recording equipment. The audio recordings were transcribed by four students supervised and checked by RO and WD. The questions posed by the manual therapists and the patients' answers were counted. The patients' demographic and clinical characteristics were also recorded. The age, gender, clinical experience, and additional educational attainment of the manual therapists were noted. To evaluate the extent of self-

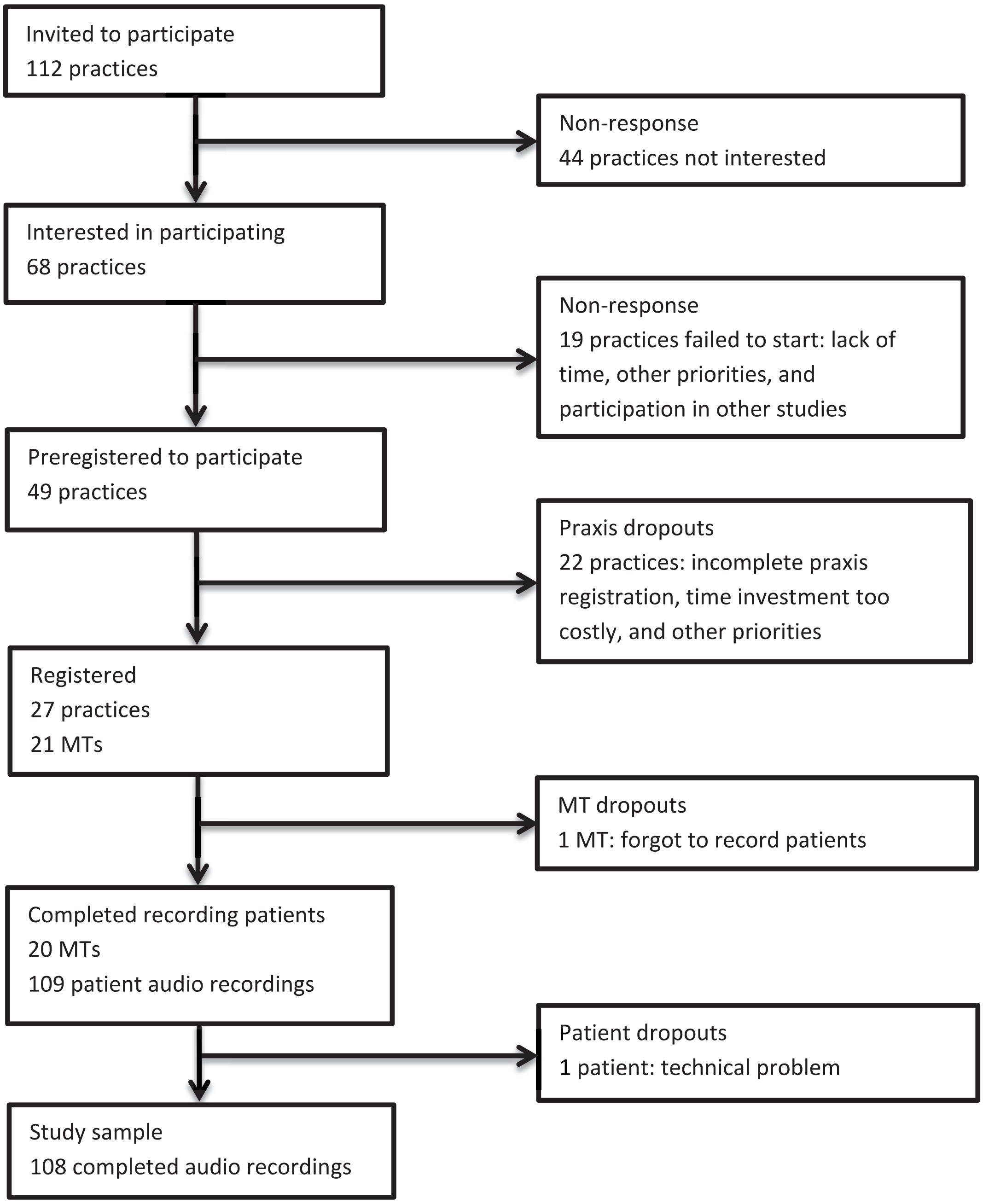

Figure. Flowchart of participating manual therapy practices and manual therapists' (MTs) responses with reasons for non-response and dropouts

estimated use of biopsychosocial history taking, the manual therapists were subsequently asked if all dimensions of the SCEBS method were dealt with.

Data analysis

The transcripts were read several times by each of the students and supervisors in order to achieve familiarity with the contents of the questions and answers during history taking. Significant phrases were identified that characterised a specific question and answer of a (sub)dimension of the SCEBS method of history taking. One point was scored for each question that adhered to (sub)dimensions of the SCEBS method.

Process indicators were scored as percentages yielding possible scores for the use of biopsychosocial history taking ranging from 0 to 100%, with the number of times an indicator was met as the numerator and the number of patients assessed as the denominator. To allow for easy interpretation, percentage scores of process indicators were categorised as negligible (0-15%), low (16-25%), weak (26-35%), very inadequate (36-45%), inadequate (46-55%), adequate (56-65%), substantial (66-75%), good (76-85%), very good (86-95%), and excellent (96-100%). The cut-off point for acceptable coverage for every dimension was set at 60%.

The estimated extent of the use of biopsychosocial history taking by the manual therapists themselves was expressed as percentages using the same categorisation as above.

Results

Phase 1: Indicator development

Sixty-eight literature-based recommendations were extracted for biopsychosocial history taking in patients with non-specific back or neck pain. After critical evaluation and checking for duplication and overlap by two members of the project group (RO and WD), the number of preselected recommendations was reduced to 51 items.

The recommendations were transformed into 51 process indicators: for instance, 'the percentage of patients who were asked about their own influence on their complaints', 'the percentage of patients who were asked about the reaction of their social environment to their complaints', or 'the percentage of patients who were asked about fear related to certain physical activities' (Appendix).

The process indicators were classified into the dimensions of the SCEBS method: Somatic dimension (n = 10), psychological dimension (Cognition n = 14; Emotion n = 6; Behaviour n = 11), and Social dimension (n = 10).

Phase 2: Indicator testing

Response rates

Of the 21 registered manual therapists, 20 (95.2%) submitted data to the pilot practice test (Figure). One hundred and nine patients participated in the study of whom one was excluded from the analysis due to a technical problem with the audio recording, leaving 108 patient recordings in the study.

Participating manual therapists and patients

The mean age of the manual therapists (n = 20) was 40.7 years (SD = 8.5) of whom 45.0% (n = 9) were female. All participants had postgraduate level education in manual therapy (Stichting Opleiding Manuele Therapie (SOMT, Educational Institute for Manual Therapy), Amersfoort, the Netherlands). The range of practice experience was eight to 22 years.

The average age of the patients (n = 108) was 42.3 years (SD = 14.1) of whom 60 (55.6%) were female. Of the 108 patients, 68 (62.9%) had back pain and 40 (37.0%) had neck pain.

Use of biopsychosocial history taking

Average percentage scores for the use of biopsychosocial history taking, according to the QIs classified into the dimensions of the SCEBS method, indicated that the extent to which the participating manual therapists met the process indicators was excellent for the Somatic dimension (98.1%), very inadequate for Cognition (42.5%) and Behaviour (37.9%), weak for Emotion (26.8%), and low for the Social dimension (17.6%) (Table). The coverage of the Somatic dimension was above the cut-off criterion of 60%.

Table. Use of biopsychosocial history taking, according to the SCEBS method: number of quality indicators (Appendix) and number and percentage scores for actual and self-estimated use by manual therapists *(n = 20)* in patients with back or neck pain *(n = 108)*

History taking	Actual use n (%)	Self-estimated use n (%)
S = Somatic dimension		
10 indicators	106 (98.1)	108 (100)
Psychological dimension		
C = Cognition		
14 indicators	46 (42.5)	65 (60.1)
E = Emotion		
6 indicators	29 (26.8)	65 (60.1)
B = Behaviour		
11 indicators	41 (37.9)	65 (60.1)
S = Social dimension		
10 indicators	19 (17.6)	43 (39.8)

Average percentage scores for the self-estimated extent of use of biopsychosocial history taking, according to the SCEBS method, by the manual therapists themselves indicated that the level of use of the Somatic dimension was excellent (100%), adequate for Cognition, Emotion, and Behaviour (60.1%) of the psychological dimension, and very inadequate for the Social dimension (39.8%) (Table).

Discussion

This study demonstrates that the use in clinical practice in manual therapy of biopsychosocial history taking in patients with back or neck pain varies widely across the various dimensions of the SCEBS method. In particular, the psychological and social dimensions of (chronic) pain were inadequately covered during history taking in these patients. Although we could not find a comparable study in the literature, these data are consistent with studies of physiotherapy care that showed poorer quality in the implementation of biopsychosocial management of musculoskeletal disorders than in the implementation of biomedical management for back and neck pain.[16] These results suggest that manual therapists involved in the primary care of patients with (chronic) musculoskeletal disorders need more in-depth training in biopsychosocial history taking, preferably adopting the SCEBS method, along with continuing education to develop and maintain skills.[17;18] With the notable exception of the somatic dimension, it is striking that the participating manual therapists overestimated their use of biopsychosocial history taking. It is possible that during the course of the patient contacts biopsychosocial information is added and subsequently integrated into the clinical reasoning processes.[17-21] A prospective study with follow-up of patient contacts could reveal the subsequent gathering of such information.

Manual therapists should be familiar not only with the biopsychosocial context of pain but also with modern insights from pain neuroscience concerning reconceptualisation of pain.[22] A sustained biomedical approach can lead to an iatrogenic effect which results in an increase in pain.[23] Although there is increasing evidence supporting the role of psychological and social factors in the emergence and persistence of chronic musculoskeletal pain, the majority of clinicians received a biomedically focused education, a focus that is also evident in the profession of manual therapy. This focus is reflected in a long tradition of treatment options based on biomechanical principles in patients with musculoskeletal disorders. This emphasis on biomedical aspects likely shapes therapists' knowledge, attitudes, beliefs, and behaviour towards (chronic) musculoskeletal pain.[22;23] In addition, the emergence of new or revised theory and subsequent changes in practice are often characterised by a significant time-lag. The integration of the biopsychosocial model into clinical practice is therefore challenging, especially for those practitioners who did not receive formal education in the application of this model in clinical assessments. The concept that (chronic) musculoskeletal pain is a condition best understood with reference to an interaction of physical (biological), psychological, and social factors is increasingly accepted in manual therapy. It is therefore not surprising that this acceptance has led to discussion of the value of manual therapy as a one-dimensional (physical) assessment in patients with back or neck pain. This has resulted in the integration of psychological and social factors in clinical practice guidelines and in multidisciplinary biopsychosocial rehabilitation programs.[24] It has also been suggested that multimodal treatments are superior to unimodal treatment (e.g., manual therapy).

Despite the development of many (theoretical) implementation strategies and activities in the field of manual therapy[25;26], programs to enhance guideline adherence including the use of standardised measurement instruments or questionnaires have so far been relatively ineffective.[27-31] It has been reported that manual therapists exhibit only moderate adherence to clinical practice guidelines and research carried out in the Netherlands has revealed that a lack of knowledge and competencies of physiotherapists with respect to the use of measurement instruments and questionnaires may hamper the implementation of guidelines.[32]

This study describes the development of QIs to measure the use of biopsychosocial history taking as a first step in clinical practice guidelines associated with (chronic) musculoskeletal pain in patients with back or neck pain. Additional evidence indicates that many interfering factors in relation to pain can only be identified by careful history taking.[33] The SCEBS method is the most commonly used method in the Netherlands for a systematic inventory and analysis of factors related to pain and this method is also integrated in the revision and actualisation of clinical practice guidelines by the Royal Dutch Society for Physical Therapy (e.g., Low Back Pain[34]).

Although no formal external validation of the set of QIs has taken place, the systematic approach and the composition of the research group underline the content validity of the QIs set derived from the SCEBS method. External validity depends on the heterogeneity of the expert panel which consisted of patient representatives, psychologists, general physicians, manual therapists, and teachers. There is a pressing need for further research in the aforementioned area that includes larger groups of both experts and patients. While in this case the response rate of participating practices was acceptable (55.1%), the relatively small self-selected sample of participating manual therapists might limit the external validity of the practice test. The manual therapists were comparable to national profiles for this group[35] and patients were comparable to participants in other Dutch studies.[27;36-38]

With a target of five patients per participant, the number of patients was adequate. The high patient response was probably due to the limited burden of recording the history taking with an audio recorder, in contrast to the greater time commitments of a randomised study or the repeated filling of questionnaires for the evaluation of treatment in clinical practice.[32] Lack of time is one of the reasons for not entering or no longer participating in clinical studies.

In addition to the years of clinical experience, the majority of participating manual therapists were also educated in the biomedical model of pain. Unsurprisingly, the use of somatic history taking was 'excellent' in this study. By contrast, the use of the psychological and social dimension was 'very inadequate' to 'low'. Unlike recent graduate manual therapists, it might have been expected that a group of manual therapists with long clinical experience would have integrated the biopsychosocial approach into their first contacts with patients. In a qualitative study, Agledahl et al[39] found that young doctors or doctors in training largely ignore the impact of symptoms on patient's daily life. This biomedical approach suggests that the next steps in the clinical reasoning process will be defined by the results of the preceding biomedical history taking.

Manual therapy is often presented as a treatment option to patients with back or neck pain within this biomedical model of pain. Traditionally, the objectives of manual therapy have been to find impairments in body functions and structures related to posture and movement which manual therapists then treat using hands-on techniques (e.g., mobilisation or manipulation of joints). In this pilot study, only data on history taking are available; no data on the remaining steps of the diagnostic and therapeutic process and the outcome of treatment were gathered. This may be regarded as a limitation of the study.

A large number of published studies and (systematic) review articles in various journals (e.g., *Pain* and *Manual Therapy*) advocate a broader view of (chronic) musculoskeletal pain.[3;5;17;20;40-42] Based on this literature and the results of our study, we urge manual therapists to make this broader vision their own. An increasing number of manual therapy curricula around the world now emphasise the biopsychosocial model in their educational programs and teach communication skills in addition to hands-on techniques.[43;44] Recent research clearly demonstrates that musculoskeletal pain is a heterogeneous condition involving biological, psychological, and social factors to varying degrees. Biopsychosocial history taking using a method such as SCEBS, in combination with the ICF and modern insights from pain neuroscience, plays a central role in the inventory of biological, psychological, and social factors and consequently in the next steps of the clinical reasoning process of manual therapists.

Conclusions and recommendations

Our results indicate that manual therapists perform screening for musculoskeletal pain mainly through the use of the somatic or biomedical dimension of (chronic) pain, according to the SCEBS method, in patients with back or neck pain. The psychological and social dimensions of chronic pain were inadequately covered by manual therapists. There is a substantial discrepancy between the actual and self-estimated use of biopsychosocial history taking. Further work should focus on the role of education of manual therapists in promoting a complete biopsychosocial history taking and follow-through within the diagnostic, therapeutic, and evaluative phases of the clinical reasoning process.

References

1. Engel GL. The biopsychosocial model and the education of health professionals. Ann N Y Acad Sci 1978;310:169-87.
2. Waddell G. 1987 Volvo award in clinical sciences. A new clinical model for the treatment of low back pain. Spine 1987;12:632-44.
3. Nijs J, Torres-Cueco R, Van Wilgen CP, Girbes EL, Struyf F, Roussel R, Van Oosterwijck J, Daenen L, Kuppens K, Vanderweeen L, Hermans L, Beckwee D, Voogt L, Clark J, Moloney N, Meeus M. Applying modern pain neuroscience in clinical practice: Criteria for the classification of central sensitization pain. Pain Physician 2014;17:447-57.
4. Roussel NA, Nijs J, Meeus M, Mylius V, Fayt C, Oostendorp R. Central sensitization and altered central pain processing in chronic low back pain: Fact or myth? Clin J Pain 2013;29:625-38.
5. Vlaeyen JW, Linton SJ. Fear-avoidance model of chronic musculoskeletal pain: 12 years on. Pain 2012;153:1144-7.
6. International Classification of Functioning, Disability and Health: ICF. Geneva, Switzerland: World Health Organization, 2001. Available at: http://www.who.int/classifications/icf/en/. (accessed 18 September, 2014)
7. Van Spaendonck KPM. Functionele klachten in de medische praktijk: Een werkmodel. Assen: Van Gorcum, 1995. [In Dutch]
8. Van Duppen D, Neirincks J, Seuntjens L. Van counselen naar cognitieve gedragstherapie. Huisarts Nu 2005;34:1-7. [In Dutch]
9. Van Spaendonck KPM, Bleijenberg G. Biopsychosociale klachten, SCEGS. Het biopsychosociale model (SCEGS). Available at: http://www.wikifysio.nl/index.php/Biopsychosociale_klachten,_SCEGS. [In Dutch] (accessed 18 September, 2014)
10. Olde Hartman TC, Blankenstein AH, Molenaar AO, Bentz van den Berg D, Van der Horst HE, Arnold IA, Burgers JS, Wiersma TI, Woutersen-Koch H. NHG-Standaard Somatisch Onvoldoende verklaarde Lichamelijke Klachten (SOLK). Huisarts Wet 2013;56:222-30. [In Dutch]

11. Hoedeman R, Wijers JHL, Van der Beek EJ, Te Koppele A. Toepassing van het SCEGS-model in de begeleiding van somatiserende werknemers. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:441-4. [In Dutch]
12. Lawrence M, Olesen F. Indicators of quality health care. Eur J Gen Pract 1997;3:103–108.
13. Grol R, Wensing M, Eccles M. Improving patient care. The implementation of change in clinical practice. Edinburgh: Elsevier Butterworth Heinemann, 2005.
14. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care 2003;15:523-30.
15. Guyatt GH, Oxman AS, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, Schünemann HJ. GRADE Working Group. Going from evidence to recommendations. BMJ 2008;336:1049-51.
16. Green AJ, Jackson DA, Klaber-Moffet JA. An observational study of physiotherapists' use of cognitive-behavioral principles in the management of patients with back and neck pain. Physiotherapy 2008;94:306-13.
17. Nijs J, Roussel N, Van Wilgen P, Köke A, Smeets R. Thinking beyond muscles and joints: Therapists' and patients' attitudes and beliefs regarding chronic musculoskeletal pain are key to applying effective treatment. Man Ther 2013;18:96-102.
18. Nijs J, Van Houdenhove B, Oostendorp RA. Recognition of central sensitization in patients with musculoskeletal pain: Application of pain neurophysiology in manual therapy practice. Man Ther 2010;15:135-41.
19. Domenech J, Sánchez-Zuriaga D, Segura-Ortí E, Espejo-Tort B, Lisón JF. Impact of biomedical and biopsychosocial training sessions on the attitudes, beliefs, and recommendations of health care providers about low back pain: A randomized clinical trial. Pain 2011;152:2557-63.
20. Van Wilgen P, Beetsma A, Neels H, Roussel N, Nijs J. Physical therapists should integrate illness perceptions in their assessment in patients with chronic musculoskeletal pain: A qualitative analysis. Man Ther 2014;19:229-34.
21. Van Trijffel E, Plochg T, Van Hartingsveld F, Lucas C, Oostendorp RA. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual physical therapy: A qualitative interview study. J Man Manip Ther 2010;18:111-8.
22. Ostelo RW, Vlaeyen JW. Attitudes and beliefs of health care providers: Extending the fear-avoidance model. Pain 2008;135:3-4.

23. Darlow B, Fullen BM, Dean S, Hurley DA, Baxter GD, Dowell A. The association between health care professional attitudes and beliefs and the attitudes and beliefs, clinical management, and outcomes of patients with low back pain: A systematic review. Eur J Pain 2012;16:3-17.
24. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RWJG, Guzman J, Van Tulder MW. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database of Systematic Reviews 2014, Issue 9, Art. No.: CD000963.
25. Jones M, Edwards I, Gifford L. Conceptual models for implementing biopsychosocial theory in clinical practice. Man Ther 2002;7:2-9.
26. Bekkering GE, Engers AJ, Wensing M, Hendriks HJ, Van Tulder MW, Oostendorp RA, Bouter LM. Development of an implementation strategy for physiotherapy guidelines on low back pain. Austr J Physiotherapy 2003;49:208-14.
27. Bekkering GE, Hendriks HJ, Van Tulder MW, Knol DL, Hoeijenbos M, Oostendorp RA, Bouter LM. Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: A cluster randomised controlled trial. Qual Saf Health Care 2005;14:107-12.
28. Rutten GM, Degen S, Hendriks EJ, Braspenning JC, Harting J, Oostendorp RA. Adherence to clinical practice guidelines for low back pain in physical therapy: Do patients benefit? Phys Ther 2010;90:1111-22.
29. Swinkels IC, Van den Ende CH, Van den Bosch W, Dekker J, Wimmers RH. Physiotherapy management of low back pain: Does practice match the Dutch guidelines? Austr J Physiotherapy 2005;51:35-41.
30. Oostendorp RA, Rutten GM, Dommerholt J, Nijhuis-van der Sanden MW, Harting J. Guideline-based development and practice test of quality indicators for physiotherapy care in patients with neck pain. J Eval Clin Pract 2013;19:1044-53.
31. Van Dulmen SA, Maas M, Staal JB, Rutten G, Kiers H, Nijhuis-van der Sanden M, Van der Wees P. Effectiveness of peer assessment for implementing a Dutch physical therapy low back pain guideline: Cluster randomized controlled trial. Phys Ther 2014;94:1396-409.
32. Swinkels RA, Van Peppen RP, Wittink H, Custers J, Beurskens AJ. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. BMC Musculoskelet Disord 2011;12:106.

33. Chodosh J, Ferrell BA, Shekelle PG, Wenger NS. Quality indicators for pain management in vulnerable elders. Ann Intern Med 2001;135:731-5.
34. Staal JB, Hendriks EJM, Heijmans M, Kiers H, Lutgers-Boomsma AM, Rutten G, Van Tulder MW, Den Boer J, Ostelo R, Custers JWH. KNGF-richtlijn Lage rugpijn. Amersfoort: Koninklijk Nederlands Genootschap voor Fysiotherapie, 2013. [In Dutch]
35. Kenens R, Hingstman L. Cijfers uit de registratie van fysiotherapeuten. Peiling 1 januari 2008. Available at: http://www.nivel.nl/beroepenindezorg/. [In Dutch] (accessed September 19, 2014)
36. Picavet HJS, Schouten JASG. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC-study. Pain 2003;102:167-78.
37. Pool JJ, Ostelo RW, Knol DL, Vlaeyen JW, Bouter LM, De Vet HC. Is a behavioral graded activity program more effective than manual therapy in patients with subacute neck pain? Results of a randomized clinical trial. Spine 2010;35:1017-24.
38. Hoving JL, De Vet HC, Koes BW, Van Mameren H, Devillé WL, Van der Windt DA, Assendelft WJ, Pool JJ, Scholten RJ, Korthals-de Bos IB, Bouter LM. Manual therapy, physical therapy, or continued care by the general practitioner for patients with neck pain: Long-term results from a pragmatic randomized clinical trial. Clin J Pain 2006;22:370-7.
39. Agledahl KM, Gulbrandsen P, Førde R, Wifstad Å. Courteous but not curious: How doctors' politeness masks their existential neglect. A qualitative study of video-recorded patient consultations. J Med Ethics 2011;37:650-4.
40. Nijs J, Meeus M, Cagnie B, Roussel NA, Dolphens M, Van Oosterwijck J, Danneels LA. Modern neuroscience approach to chronic spinal pain: Combining pain neuroscience education with cognition-targeted motor control training. Phys Ther 2014;94:730-8.
41. Nijs J, Van Wilgen CP, Van Oosterwijck J, Van Ittersum M, Meeus M. How to explain central sensitization to patients with 'unexplained' chronic musculoskeletal pain: Practice guidelines. Man Ther 2011;16:413-8.
42. Meeus M, Nijs J, Van Oosterwijck J, Van Alsenoy V, Truijen S. Pain physiology education improves pain beliefs in patients with chronic fatigue syndrome compared with pacing and self-management education: A double-blind randomized controlled trial. Arch Phys Med Rehabil 2010;91:1153-9.

43. Vissers D, Daele UV, De Hertogh W, De Meulenaere A, Denekens J. Introducing competency-based education based on the roles that physiotherapists fulfil. J Nov Physiother Phys Rehabil 2014;1:110.
44. Lluch Girbés E, Meeus M, Baert I, Nijs J. Balancing "hands-on" with "hands-off" physical therapy interventions for the treatment of central sensitization pain in osteoarthritis. Man Ther 2015;20:349-52.

Appendix

SCEBS method (Dutch: SCEGS methode)

S = Somatic dimension (Dutch: Somatische dimensie)

1. What are your complaints?
2. When did the complaints begin?
3. What is the nature, the location, the intensity of the complaints?
4. How often do the symptoms occur?
5. How long do the symptoms last?
6. Have you had these symptoms before?
7. Can you move your back/neck?
8. Have you experienced any stiffness?
9. What do the X-ray results show?
10. What do laboratory tests show?

Psychological dimension (Dutch: Psychologische dimensie)

***C** = **C**ognition (Dutch: **C**ognitie)*

Expectations

11. What do you expect from me?
12. What do you think I can do for you?

Explanations (attribution)

13. What do you think yourself?
14. Do you yourself have any explanation for your complaints?
15. Do you sometimes think "if it isn't this or that"?

Thinking about complaints/thinking that worsens complaints (catastrophising)

16. How do you feel when you have symptoms?
17. What do you think at that moment?
18. How do you react?

Ideas about personal influence on complaints (self-efficacy)

19. Do you personally have any influence on the complaint?

20. Can you positively influence the complaint?

21. If so, how?

22. Is there anything you yourself can do to reduce your complaint?

23. Do complaints resolve more quickly when you rest?

24. Do complaints lessen when you think about something or someone else?

***E** = **E**motion (Dutch: **E**motie)*

25. Given that you have these complaints, how do you feel about it?

26. Do the complaints disturb your emotional balance?

27. Are you insecure?

28. Are you depressed?

29. Are you anxious?

30. Do you ever feel overwhelmed by the complaints?

***B** = **Behaviour** [Dutch: **G**edrag]*

Dealing with the complaint

31. What do you do if you have symptoms?

32. What do you do to reduce symptoms?

33. To what extent is this successful?

Limitations to activities

34. Which activities are hindered by your complaints?

35. To what extent?

Avoidance

36. What don't you do or no longer do when you have symptoms?

37. Since when?

38. Are you anxious about particular activities?

39. What do other people notice about your behaviour when you have symptoms?

Talking about complaints

40. Do you talk about your complaints? Who with? How often?
41. What do you tell them?

S = Social dimension (Dutch: Sociale dimensie)

42. Do the people around you notice when you have complaints?
43. What do they notice?
44. How do you react to your complaints?
45. What do the people around you think about your complaints?
46. How do the people around you react to your complaints?
47. Where does your partner think that your complaints come from?
48. How did the people around you react when you told them what the doctor said?
49. How do you now feel about this?
50. Do the complaints affect your social life?
51. Did you need to adapt your work/hobby/sport to your complaints?

Chapter 6

Indicating spinal joint mobilisations and manipulations in patients with neck or low-back pain: Protocol of an inter-examiner reliability study among manual therapists

Emiel van Trijffel, Robert Lindeboom, Patrick Bossuyt, Maarten Schmitt, Cees Lucas, Bart Koes, Rob Oostendorp

Chiropr Man Therap 2014;22:22

Abstract

Background: Manual spinal joint mobilisations and manipulations are widely used treatments in patients with neck or low-back pain. Inter-examiner reliability of passive intervertebral motion assessment of the cervical and lumbar spine, perceived as important for indicating these interventions, is poor within a univariable approach. The diagnostic process as a whole in clinical practice in manual therapy has a multivariable character, however, in which the use and interpretation of passive intervertebral motion assessment depend on earlier results from the diagnostic process. To date, the inter-examiner reliability among manual therapists of a multivariable diagnostic decision-making process in patients with neck or low-back pain is unknown.

Methods: This study will be conducted as a repeated-measures design in which 14 pairs of manual therapists independently examine a consecutive series of a planned total of 165 patients with neck or low-back pain presenting in primary care physiotherapy. Primary outcome measure is therapists' decision about whether or not manual spinal joint mobilisations or manipulations, or both, are indicated in each patient, in isolation or as part of a multimodal treatment. Therapists will largely be free to conduct the full diagnostic process based on their formulated examination objectives. For each pair of therapists, 2x2 tables will be constructed and reliability for the dichotomous decision will be expressed using Cohen's kappa. In addition, observed agreement, prevalence of positive decisions, prevalence index, bias index, and specific agreement in positive and negative decisions will be calculated. Univariable logistic regression analysis of concordant decisions will be performed to explore which demographic, professional, or clinical factors contributed to reliability.

Discussion: This study will provide an estimate of the inter-examiner reliability among manual therapists of indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain based on a multivariable diagnostic reasoning and decision-making process, as opposed to reliability of individual tests. As such, it is proposed as an initial step towards the development of an alternative approach to current classification systems and prediction rules for identifying those patients with spinal disorders that may show a better

response to manual therapy which can be incorporated in randomised controlled trials. Potential methodological limitations of this study are discussed.

Introduction

Neck and low-back pain are common and costly disorders in adult general populations.[1-6] Manual spinal joint mobilisations and manipulations are widely used treatments in patients with these complaints.[7;8] Although the underlying mechanisms of these treatments are far from understood, spinal joint mobilisations and manipulations are effective as well as cost-effective in patients with non-specific neck and low-back pain although no more effective than other treatment modalities.[9-14]

Traditionally, manual therapy has a strong focus on the diagnostics, treatment, and evaluation of spinal joint function by emphasising the use of passive physiological and accessory movements.[15-17] Passive intervertebral motion (PIVM) assessment is used to judge the quantity and quality of functions of spinal motion segments and is assumed to play an important role in diagnostically classifying patients and selecting treatment.[18] Dutch, New Zealand, and USA manual therapists indeed believe that passive spinal mobility testing is important for deciding on manual mobilisation or manipulation as a treatment option.[19;20] Moreover, a recent international, multidisciplinary survey showed that PIVM assessment is the most commonly used impairment outcome measure in patients with neck pain.[21]

In order to yield accurate and uniform decisions about treatment options for patients, test results need to be reliable.[22] Reliability is a component of reproducibility along with agreement and reflects the extent to which test results can diagnostically discriminate between patients despite measurement errors.[23;24] Agreement, on the other hand, concerns the possibility of examiners to obtain the same test results on different measurement occasions.[25] Systematic reviews have consistently shown poor inter-examiner reliability for spinal physical tests, and for PIVM assessment in particular.[26-30] However, the large majority of studies investigating the reliability of physical tests and PIVM assessment can be regarded as test research following a single-test or univariable approach thus neglecting the multivariable character of the diagnostic process as opposed to diagnostic research.[31]

Physiotherapists conduct a diagnostic process by collecting data through interview and physical examination and by generating hypotheses as to why a problem exists in order to reach a decision about appropriate patient management.[32;33] During this diagnostic process, manual therapists indeed seem to apply, amongst others, a hypothetico-deductive way of clinical reasoning.[34;35] PIVM assessment is usually conducted after history taking, questionnaires, and other physical tests and is indicated after interpreting earlier clinical information and formulating specific hypotheses about spinal joint dysfunction.[35] Moreover, Canadian manual therapists reported to decide on manual mobilisation or manipulation based on their whole clinical assessment and clinical reasoning in a patient.[36] It is therefore reasonable to assume that the diagnostic process in manual therapy has a multivariable character.

Over the last three decades, many systems have been developed for classifying patients with spinal disorders, in particular for those with low-back pain.[37] A systematic review found 28 systems for classifying chronic low-back pain alone and it was concluded that there was insufficient evidence to support or recommend any particular system for use in clinical description, determining prognosis, or predicting response to treatment.[38] Some systems were tested for their inter-examiner reliability, but the evidence was either conflicting or moderate to strong for poor reliability.[27] On the other hand, using clusters of tests for diagnosing sacroiliac joint dysfunction yielded acceptable reliability.[39-41] However, the majority of these systems either lack evidence for their reliability, only use certain parts of the clinical examination (e.g., only physical tests), are prescriptive in their application, do not include PIVM assessment, are not related to manual therapy interventions, or do not direct towards treatment decisions. Some systems[42;43] were developed as treatment-based classification algorithms for subgrouping patients with low-back pain and were strongly based on factors derived from several clinical prediction rules.[44-47] However, these rules lack validation, and methodological and statistical issues regarding their development have been raised.[48] In contrast to the field of classification systems for low-back pain, the development and number of systems for classifying neck pain patients lie far behind. Besides a treatment-based classification system for physiotherapy interventions[49], clinical prediction rules have been derived to identify factors that predict response to spinal manipulation in patients with neck pain but with identical problems as in the rules for low-back pain as mentioned above.[50-55] In a systematic review, Gemmell & Miller[56] found poor inter-examiner reliability of multitest regimens using only physical tests for identifying manipulable spinal lesions in chiropractic. Including pain scores and medical history next to manual examination of spinal motion segments resulted in high accuracy in identifying neck pain patients.[57] To summarise, however, the value of the diagnostic process as a whole to classify patients with neck or low-back pain in order to decide whether or not spinal mobilisations or manipulations are indicated remains unclear.

This is the protocol of a study that aims to determine the inter-examiner reliability among Dutch manual therapists of indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain based on a multivariable, hypothesis-based diagnostic reasoning and decision-making process. Secondly, using univariable logistic regression analysis of concordant decisions about indications, we will explore which demographic, professional, and clinical factors can explain variation in reliability of therapists' decisions with specific attention to the contribution of PIVM assessment.

Methods

Design

This study will be conducted as a repeated-measures design in which pairs of manual therapists independently examine a consecutive series of patients with neck or low-back pain presenting in primary care physiotherapy in the Netherlands. Primary outcome measure is therapists' decision about whether or not spinal manual therapy (SMT) is indicated in each patient, in isolation or as part of a multimodal treatment. SMT is defined here as either spinal joint mobilisations or manipulations, or both. Therapists will largely be free to conduct the full diagnostic process as they are routinely used to.

Participants

Consecutive patients aged 18 years or older presenting with a primary complaint of neck or low-back pain, either referred to primary care physiotherapy by their general practitioner or medical specialist, or by self-referral, will be eligible for participation in the study. Neck pain is defined as pain in the region between the superior nuchal line, the external occipital protuberance, the spines of the scapula, the superior border of the clavicula, and the suprasternal notch, with or without radiation to the head, trunk, or upper limbs.[58] Patients will not be eligible when headache or dizziness is their dominant complaint. Low-back pain is defined as pain or discomfort localised below the costal margin and above the inferior gluteal folds, with or without radiation to the lower limbs.[59] All patients who are assumed to have non-specific or (non-serious) specific neck or low-back pain with a potential indication for SMT will be included. Patients who are not able to speak or read Dutch fluently will be ineligible. Patients will receive verbal and written information on all aspects of the study and will be asked to provide written consent at their inclusion. The Central Committee for Research involving Human Subjects (CCMO, the Hague, the Netherlands) decided that a full evaluation of the study protocol by a medical ethical committee was not required because patients will undergo a diagnostic process similar to routine clinical practice.

Examiners

Examiners will be manual therapists working at least 20 hours a week in their private practices in the Netherlands and registered by the Dutch Association for Manual Therapy or the Royal Dutch Society for Physical Therapy. From a database of those graduated from the

Institute for Master Education in Musculoskeletal Therapy (SOMT: Stichting Opleidingen Musculoskeletale Therapie, Amersfoort, the Netherlands), 14 pairs of manual therapists will be invited to participate. Each pair works together in the same practice and practices will be selected based on their ability to logistically organise the study. We aim to include therapists who vary in years of clinical experience in manual therapy. Therapists will attend an information session followed by a two-hour training session in which procedures for digitally registering data are explained and practised. They will not receive additional training in history taking, physical examination procedures, or using questionnaires. Pairs of therapists will be strictly requested not to discuss their experiences during the study with each other until their last patient has been included. Gender, age, years of clinical experience in manual therapy, highest diploma, practice setting, weekly amount of work related to spinal disorders (hours), teaching experience (yes/no), and participation in research (yes/no) will be recorded as professional characteristics from the participating therapists.

In each practice, a third colleague will function as a research assistant to coordinate the inclusion and flow of patients. Research assistants will be instructed with respect to applying the inclusion criteria, the order of assigning patients to therapists, and assuring blinding procedures.

Procedures

From eligible patients, demographic (gender, age, marital status, working status) and clinical (type of complaints (neck or low-back pain), duration of complaints (days), radiation (yes/no), traumatic origin (yes/no), comorbidity (yes/no)) data will be recorded as baseline data by the local research assistant. In addition, baseline pain and disability will be determined using the Numeric Pain Rating Scale (NPRS 0-10, higher scores indicate higher pain intensity), and the Quebec Back Pain Disability Scale (QBPDS 0-100, higher scores indicate higher disability) for low-back pain patients and the Neck Disability Index Dutch Language Version (NDI-DLV 0-50, higher scores indicate higher pain and disability) for neck pain patients, respectively. The NPRS is a reliable and valid scale to measure pain intensity in adults.[60] The Dutch version of the QBPDS is a reliable and valid instrument for measuring disability in low-back pain patients[61] and the Dutch version of the NDI is recommended for measuring pain and disability in patients with neck pain.[62]

All baseline data will be available to each therapist before he or she starts the diagnostic process. The first therapist of each pair will be the treating therapist to whom the patient was assigned to, so the order in which both therapists act as the first examiner will vary according to the practice's planning. The first therapist will screen all consecutive patients with neck or low-back pain for the presence of red flags.[63] In accordance with guidelines in the Netherlands[64], patients suspected of having serious (spinal or non-spinal) pathology will not enter the study which will be recorded. Patients will then undergo a full history taking by the first therapist. The therapist will record his or her findings as well as proposed hypotheses about patient's health status by formulating explicit objectives for further examination. The therapist will then choose the diagnostic procedures (e.g. observation, physical tests, performance tests, questionnaires) that he or she plans to perform in the patient. After performing each procedure, its outcome will be recorded. If PIVM assessment is indicated, therapists will use three-dimensional coupled movements in flexion and extension directions for each individual motion segment.[65] Movements will be judged on mobility (hypermobile-normal-hypomobile), resistance perceived by the therapist during the movement (increased resistance or stiffness yes/no), resistance perceived by the therapist at the end of the movement (or end-feel) (increased resistance or stiffness at the end of the movement yes/no), and pain provocation (yes/no). Therapists will perform a maximum of three repetitions for each movement per direction per spinal motion segment to afford the best stiffness discriminability.[66]

The therapist will then be asked to record whether he or she has made any changes to the original examination objectives as well as to specify these changes, and a diagnostic conclusion in terms of specific or non-specific neck or low-back pain is given. Finally, the therapist will make the decision about whether or not SMT is indicated in the patient and, when indicated, it will also be stated whether mobilisations or manipulations, or both, are indicated, and to which spinal motion segments these techniques would be targeted. In addition, the therapist will rate his or her level of certainty of the primary decision about the indication on a bipolar seven-point scale ranging from -3 (completely uncertain) to 3 (completely certain). It will also be recorded which other interventions he or she believes would further be indicated in the patient. However, at this point, no actual treatment will be provided.

After the first therapist has performed the full examination, he or she will leave the examination room and the patient will be given a 10-minute break. After checking whether all data have been registered, the research assistant then guides the second therapist into the room and makes sure that there is no visual or verbal contact between the two therapists. The second therapist will then conduct the full diagnostic process, excluding the screening for red flags, whilst being unaware of the outcomes of the first examination. Patients will be requested not to mention any outcomes or conclusions from the first examination. Both therapists will record all their findings and data into a fit-for-purpose software program. The research assistant will check whether all data have been entered by both therapists.

Statistical analysis

Demographic and clinical baseline characteristics of patients will be summarised using descriptive statistics. Absolute and relative frequencies are used to describe categorical data. Ordinal data relating to patients' pain and disability will be described with their median and interquartile range. Normally distributed numerical data will be summarised by their mean and standard deviation. In case of non-normality, median and interquartile range are presented. Examination objectives as formulated by therapists will be classified by one researcher (EvT) according to the framework of the World Health Organization's *International Classification of Functioning, Disability and Health* (ICF)[67] to describe patients' functioning in terms of impairments of neuromusculoskeletal and movement-related functions, activity limitations, and participation restrictions, and personal and environmental factors. Diagnostic procedures will be listed and described with their frequencies, and also outcomes of PIVM assessment, changes to the original examination objectives, diagnostic conclusions, and examiners' level of certainty of their decision about the treatment indication will be summarised. Concordance between the formulated examination objectives concerning spinal joint motion function and the actual use of PIVM assessment will be presented as frequencies.

For each pair of therapists, 2x2 tables will be constructed and reliability for the dichotomous positive or negative decisions about whether or not SMT is indicated will be calculated as chance-corrected reliability using Cohen's kappa.[68] As recommended by Cicchetti & Feinstein[69] and Byrt et al[70], observed agreement (%), prevalence of positive decisions (mobilisations and/or manipulations indicated) relative to the total number of indications, prevalence index (PI), bias index (BI), and specific agreement (%) in positive (p_{pos}) and negative (p_{neg}) decisions will be calculated in order to evaluate whether kappa was influenced by high prevalence of positive or negative decisions, or by systematic bias between examiners. PI reflects the difference between the proportion of agreement on positive indications as compared to that of negative indications. PI ranges between 0 and 1, and is high when the prevalence of concordant positive (or negative) indications is high, chance agreement is consequently also high, and kappa is reduced accordingly (prevalence effect).[71] BI provides a quantification of the extent to which examiners disagree on the proportions of positive (or negative) indications. BI also ranges between 0 and 1, and is high when the difference between the discordant indications is high, chance agreement is consequently low, and kappa is inflated accordingly (bias).[71] P_{pos} and P_{neg} are the proportions of agreement on positive and negative indications, respectively, relative to the total number of positive and negative indications, respectively, from both therapists. Overall kappa (95% CI) will be calculated as a generalised chance-corrected reliability across all pairs of therapists. See Appendix for formulas.

In addition, for each pair of therapists, separate 2x2 tables will be presented for judgements about the indication for PIVM assessment and for judgements about mobility, end-feel, and pain provocation obtained from PIVM assessment (four tables in total). Observed agreement, prevalence of positive decisions, PI, BI, p_{pos}, p_{neg}, and overall kappa (95% CI) will also be calculated. Analyses will be conducted using DAG_Stat.[72]

Kappa (95% CI) is interpreted in accordance with value labels as assigned by Landis & Koch[73]: <0.00: poor, 0.00-0.20: slight, 0.21-0.40: fair, 0.41-0.60: moderate, 0.61-0.80: substantial, 0.81-1.00: almost perfect. We arbitrarily assume a lower bound of the 95% CI around overall kappa of 0.60 to indicate acceptable reliability.

Univariable logistic regression analysis will be performed to explore which demographic, professional, and clinical factors contributed to the reliability of therapists' decision-making. Firstly, patients' demographic and clinical factors at baseline will concern their gender, age, type of complaints, duration of complaints (less or more than three months), radiation, traumatic origin, comorbidity, pain intensity, and disability. Such factors are associated with variation in diagnostic accuracy[74], but evidence in the context of reliability is lacking. Secondly, therapists' professional factors will include their clinical experience and weekly amount of work related to spinal disorders. Weekly amount of work related to spinal disorders was positively associated with perceived importance and confidence related to the use and interpretation of PIVM assessment[20] and may, therefore, contribute to variation in diagnostic decision-making. In addition, other clinical factors will be explored involving PIVM assessment (indicated or not, and judgements on mobility, resistance, and pain provocation), the diagnostic conclusion (specific or non-specific neck or low-back pain), therapists' level of certainty of their decision about the treatment indication, and the concordance between examination objectives and the use of PIVM assessment. Factors will be entered in the model as single covariates with the concordant decisions, either positive or negative, as the dependent variable. Concordant decisions will be coded as 1 while the discordant decisions will be coded 0. Therapists' experience and work related to spinal disorders will be entered as mean scores from each pair. A p-value <0.05 indicates a statistically significant association between a factor and a concordant decision about whether or not SMT is indicated.

With a sample size of 165, a two-sided 95% CI around kappa would extend ±0.109 from the observed value of kappa, assuming a true value of kappa of 0.70, and a prevalence of positive decisions of 50%. Consequently, each pair of examiners will be asked to include 12 patients. Multiple imputation will be used to handle records with data points missing at random. If, for any reason, data on the primary outcome measure are not available or obtainable from one or both therapists, data from this patient will be excluded from the analysis and the pair of therapists will be asked to include a new patient. Analyses will be conducted using IBM SPSS Statistics for Windows version 22.

Discussion

The results of this study will provide (1) an estimate of the inter-examiner reliability among manual therapists of indicating SMT in patients with neck or low-back pain based on a multivariable diagnostic reasoning and decision-making process, as opposed to the reliability of individual clinical tests, and (2) a first exploration of which demographic, professional, or clinical factors can explain variation in the reliability of therapists' decision-making with specific attention to the contribution of PIVM assessment. We do not aim or hypothesise that reliability from a multivariable approach to clinical diagnostics will be higher than that from individual test diagnostics. Rather, we believe that such an estimate will be a more real resemblance of the reliability among therapists of making decisions in clinical practice concerning the distinction between patients who are indicated for SMT and those who are not. In addition, this approach will add to the ongoing discussion of the identification of specific subgroups of patients that may be more likely to respond to SMT and we propose alternative research strategies for establishing treatment effects.

It has been recognised that treatment effects of SMT, or any other physiotherapy modality for that matter, especially in patients with low-back pain, are, on average, small which may be due to heterogeneity of patients obscuring a wide range of individual treatment responses and variation of treatment effects.[75] Ever since the mid-nineties of the last century, identifying subgroups of patients that may benefit more from specific or targeted interventions has had the highest research priority.[76-81] As a result, there has been a proliferation of subgrouping systems aiming to identify people with a particular pathoanatomical condition, a particular prognosis, or those that are more likely to respond favourably to treatment.[82] Primary care clinicians themselves do not believe that low-back pain is one condition and they treat patients differently based on patterns of clinical signs and symptoms.[83] Moreover, they classify patients predominantly based on pathoanatomy, but they show little consensus regarding these related patterns.[84] With the aim to identify patients that may be more likely to show a positive response to spinal manipulation, clinical prediction rules have been derived to identify predictors in patients with neck or low-back pain.[44-47,50-55] Unfortunately, systematic reviews have consistently concluded that there is, as yet, insufficient evidence to support the general application of these rules.[85-89] Another systematic review found significant treatment effects favouring subgroup-specific SMT over a number of comparison treatments for pain and disability at short and intermediate follow-up based on low-quality trials.[90] Foster et al[75] concluded that no subgrouping approaches have yet passed the tests for clinical value and robustness of evidence, and there is still a long way to go before closer matching of treatments to patient characteristics becomes a clinical reality. Indeed, two decades after the derivation of the Ottawa Ankle Rules[91], their validation and implementation is still an ongoing research process worldwide and it can be assumed that following a similar pathway for far more complex problems such as the treatment of non-specific neck and low-back pain may be even more time-consuming.

When determining treatment effects of SMT, randomised controlled trials currently do not make use of patients' full clinical health profile according to the domains of the ICF for targeting treatment. For instance, Cochrane Reviews consider primary studies including participants only based on their age and the presence of pain with or without radiation.[11;13;14] The resulting heterogeneity among trial participants and the subsequent dilution of treatment effects may be deleterious to SMT as its effectiveness may be underestimated for certain groups of patients. The majority of primary studies in patients with neck pain do not apply well-defined clinical criteria to select patients for SMT and if they do, they use only one physical test such as a mobility test or a pain provocation test in order to diagnose neck pain from a mechanical origin.[92] It is stated that clinical tests are not valid or reliable to allow targeting treatment in clinical trials.[84] This is certainly true when the reliability of individual physical tests is considered.[26-30] However, several of the increasingly popular predication rules also contain clinical variables that are unreliable including PIVM assessment.[42;46;88] Targeting SMT to a more homogeneous group of patients with neck or low-back pain based on a multivariable diagnostic process resembling clinical practice may outweigh the disadvantages of the current selection procedures in randomised controlled trials.

Awaiting evidence from the further validation of prediction rules and other classification systems, our study could offer an initial step towards a faster and easier development of an alternative approach to the identification of those patients with spinal disorders that may show a better response to SMT based on a multivariable decision process. A satisfactory level of reliability is a prerequisite for incorporating such decision-making into the design of randomised controlled trials for establishing treatment effects of SMT and thereby validating the approach. When reliability (lower bound of 95% CI around kappa) exceeds 0.60 and with BI, arbitrarily, <0.10, patients with neck or low-back pain with a positive indication can be randomised to receive, for instance, either manual mobilisations or manipulations, or both, within a multimodal treatment on the one hand or multimodal treatment without mobilisations or manipulations on the other (Figure A). Should reliability be below this cut-off but with p_{pos} (or p_{neg}), arbitrarily, >60%, this strategy can still be used by randomising only those patients of which the indication was agreed upon by two manual therapists (Figure B). P_{pos} and p_{neg} here indicate the absolute specific agreement on positive or negative indications, respectively, between therapists.[25]

With respect to our second research objective, it is important to note that empirical evidence for sources of bias and variation in reliability studies is lacking contrary to studies of diagnostic accuracy.[74;93-95] Variation arises from differences between studies, for example, in terms of demographic and disease features of study participants, characteristics of examiners, setting, or test protocol. As such, it does not lead to biased estimates of reliability, but it can limit the applicability of study results.[94] Knowledge of factors that explain variation in reliability may inform ways to improve reliability. For instance, examiner training and choosing a group of more heterogeneous study participants have been mentioned as improvement strategies, but both have their limitations and lack supporting evidence.[24] Systematic reviews may reveal subgroups of participants, examiners, or tests that consistently show higher or lower reliability. In systematic reviews, between-study comparisons are conducted to search for these subgroups as sources of variation. However, these comparisons are less valid as they are hampered by the often strong clinical and methodological heterogeneity between studies.[96] In addition, the identification of these sources of variation becomes even more troublesome when reliability is consistently low (or high) across studies. Within-study comparisons are the preferred method to explore variation in reliability. To date, very few studies have been undertaken in the field of manual therapy with this aim and method. Cook et al[97] investigated factors related to the large variability of forces used during passive accessory intervertebral movements and they found that examiners’ age, gender, experience, background and education, and frequency of use did not contribute to this variation. We present simple logistic regression analysis of concordant decisions as a flexible method that can easily be incorporated in any reliability

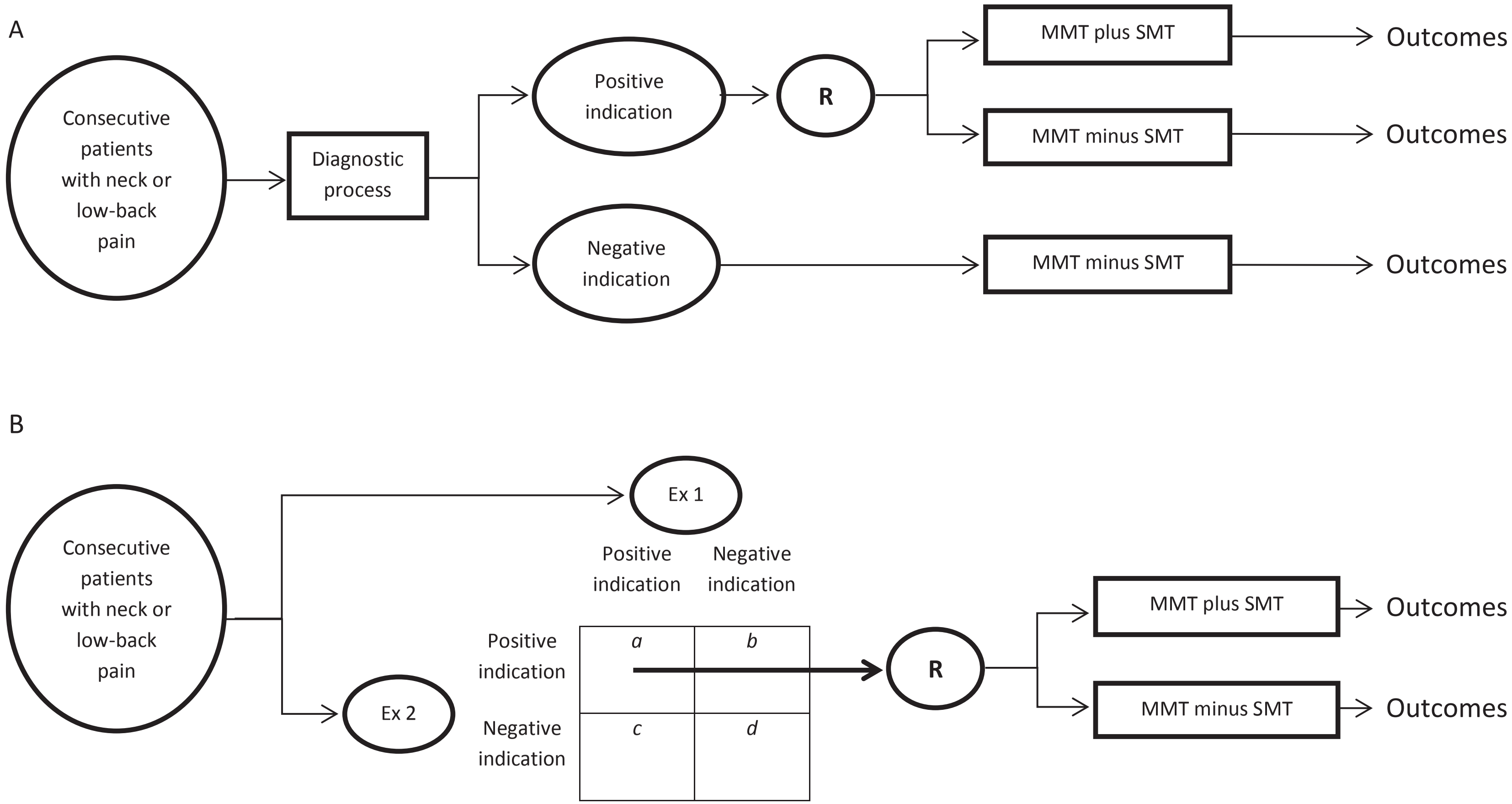

Figure. A. Design of an RCT including patients positively indicated for SMT when lower bound of 95% CI around kappa >0.60 and BI <0.10
B. Design of an RCT including only patients positively indicated for SMT by two examiners when kappa <0.60 but p_{pos} (or p_{neg}) >60%.

BI: bias index, Ex: examiner, MMT: multimodal treatment, SMT: spinal manual therapy (spinal mobilisation or manipulation, or both), R: randomisation, RCT: randomised controlled trial study to explore and explain variation in reliability from a large number of demographic, professional, and clinical factors.

Potential limitations of this study

This study protocol presents several new approaches to investigating and analysing decision-making in manual therapy and to reliability research in general. Several of its methods need further discussion in order to appraise their effect on the validity and generalisability of the study's results. First, establishing examination objectives for physical examination by physiotherapists has been used in earlier studies.[98;99] However, the prospective formulation and registration of examination objectives is far from common practice for physiotherapists in the Netherlands.[100] The specific training of our examiners in the formulation and digital registration of these objectives may diminish the generalisability of the estimated reliability of indicating SMT. We encourage that establishing and prospectively registering of examination objectives become an integral part of clinical practice in physiotherapy.

Stability of participants' characteristics is a prerequisite for the valid estimation of reliability.[23] However, very few empirical data are available as to the minimum length of the time period between test procedures that ensures that patients' responses to questions and physical tests such as joint motion assessment will remain unchanged. Shirley et al[101] reported that stiffness responses to repeated mechanical postero-anterior loading of lumbar motion segments returned to the pre-testing state within five minutes. On the other hand, a 30-minute recovery period after 30 minutes of *in vitro* creep loading of the lumbar spine was not sufficient to return to the baseline situation.[102] By incorporating a 10-minute break for patients between examinations and limiting the number of movement repetitions during PIVM assessment, we are more confident that underestimation of reliability will be avoided. Research into the natural variation over time within and between individuals regarding joint mobility and other body functions, as well as into the variation induced by the physical examination itself, is needed.

Our sample size calculation strongly depends on the assumed prevalence of positive indications which was based on data from the numerous studies on practice patterns among physiotherapists in the treatment of patients with neck or low back pain.[103-113] Within the large variation in choices of treatment options by therapists, mobilisations and manipulations were only rarely among the most preferred options and their frequency of use ranged from 16% to 83% and from 2% to 37%, respectively. These figures were not substantially different for specific subgroups of manual therapists who reported remarkably low frequencies of use of manipulations in the cervical region.[36;114-116] As we will consider reliability of indicating either mobilisations or manipulations, or both, we assume a 50% prevalence of positive indications. Choosing a higher or lower prevalence would have resulted in a larger required sample.[117]

In our sample of manual therapists and patients, we cannot rule out the possibility of a substantially higher (or lower) prevalence of positive indications for SMT. Because of such a skewed distribution of decisions, a distorted interpretation of kappa could then occur. Recently, kappa, as a relative measure of reliability, has been criticised because it can only provide information about the ability to distinguish between patients on a sample level.[25] The authors suggest using the specific agreement parameters (p_{pos} and p_{neg}) as absolute measures to quantify observer variation regarding a certain diagnosis or decision on an individual patient level.[25] No single omnibus index, however, can be satisfactory for all purposes and situations.[69;70] Therefore, we will calculate all recommended parameters from the 2x2 tables to allow full interpretation of reliability and agreement as related to the prevalence of concordant and discordant indications. We will not, however, correct kappa for prevalence effects and bias, for instance by calculating prevalence-adjusted bias-adjusted kappa, because this would generate values of reliability that no longer relate to the original situation.[117;118]

We will select pairs of manual therapists as examiners that share a common educational background. With this background from the largest institute for manual therapy education in the Netherlands, they likely form a representative sample from the Dutch population of manual therapists registered with the Dutch Association for Manual Therapy or the Royal Dutch Society for Physical Therapy. Manual therapy education in the Netherlands is strongly embedded within international concepts. In these traditional concepts, especially passive joint motion assessment takes a prominent place.[15] Therefore, we suppose that the results of this study will to a certain extent be generalisable to populations of manual therapists outside the Netherlands. We do, however, suggest that this study be replicated over different countries and concepts to account for local idiosyncrasies in clinical reasoning and decision-making. In addition, for practical reasons, we will choose pairs of manual therapists that work in the same practice. This may inflate reliability and by pairing therapists with different levels of experience, we aim to minimise this potential threat to the validity of the study.

Finally, when analysing the reliability of indicating SMT, we will not distinguish specifically between mobilisations or manipulations. Despite the disparate mechanisms of these interventions[9;119], no evidence is available on whether one or the other, or both, should be preferred in any clinical situation. Results of randomised controlled trials have been conflicting so far.[120-123] New research should focus on the relationship between clinical findings, the choice for either mobilisation or manipulation, and subsequent clinical outcomes.

References

1. Borghouts JAJ, Koes BW, Bouter LM. Cost-of-illness in neck pain in the Netherlands in 1996. Pain 1999;80:629-36.
2. Côte P, Cassidy D. Carroll L. The Saskatchewan health and back pain survey. The prevalence of neck pain and related disability in Saskatchewan adults. Spine 1998;23:1689-98.
3. Hogg-Johnson S, Van der Velde G, Carroll LJ, Holm LW, Cassidy JD, Guzman J, Côte P, Haldeman S, Ammendolia C, Carragee E, Hurwitz E, Nordin M, Peloso P. The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine 2008;33:S39-51.
4. Linton SJ, Hellsing AL, Hallden K. A population-based study of spinal pain among 35-45 year old individuals. Prevalence, sick leave and health care use. Spine 1998;23:1457-63.
5. Van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. Pain 1995;62:233-40.
6. Waddell G. Low back pain: a twentieth century health care enigma. Spine 1996;21:2820-5.
7. Assendelft WJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. Ann Intern Med 2003;138:871-81.
8. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Brønfort G, Hoving JL. Manipulation or mobilization for neck pain: a Cochrane Review. Man Ther 2010;15:315-33.
9. Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: A comprehensive model. Man Ther 2009;14:531-8.
10. Bronfort G, Haas M, Evans R, Leininger B, Triano J. Effectiveness of manual therapies: The UK evidence report. Chiropr Osteopat 2010;18:3.
11. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Brønfort G, Hoving JL. Manipulation or mobilisation for neck pain. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD004249.

12. Michaleff ZA, Lin C-WC, Maher CG, Van Tulder MW. Spinal manipulation epidemiology: Systematic review of cost-effectiveness studies. J Electromyogr Kinesiol 2012;22:655-62.
13. Rubinstein SM, Van Middelkoop M, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for chronic low-back pain. Cochrane Database of Systematic Reviews 2011, Issue 2. Art. No.: CD008112.
14. Rubinstein SM, Terwee CB, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for acute low-back pain. Cochrane Database of Systematic Reviews 2012, Issue 9. Art. No.: CD008880.
15. Farrell JP, Jensen GM. Manual therapy: A critical assessment of role in the profession of physical therapy. Phys Ther 1992;72:843-52.
16. Maher C, Latimer J. Pain or resistance – the manual therapists' dilemma. Austr J Physiotherapy 1992;38:257-60.
17. Van Ravensberg CDD, Oostendorp RAB, Van Berkel LM, Scholten-Peeters GGM, Pool JJM, Swinkels RAHM, Huijbregts PA. Physical therapy and manual physical therapy: Differences in patient characteristics. J Man Manip Ther 2005;13:113-24.
18. Jull G, Treleaven J, Versace G. Manual examination: Is pain provocation a major cue for spinal dysfunction? Austr J Physiotherapy 1994;40:159-65.
19. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.
20. Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine. A survey of Dutch physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51.
21. MacDermid JC, Walton DM, Côté P, Lina Santaguida P, Gross A, Carlesso L. Use of outcome measures in managing neck pain: An international multidisciplinary survey. Open Orthop J 2013;7:440-60.
22. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307-17.
23. De Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
24. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use (4^{th} edition). Oxford, UK: Oxford University Press, 2008.

25. De Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's κ. BMJ 2013;346:f2125.
26. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
27. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Austr J Physiother 2006;52:91-102.
28. Seffinger MA, Najm WI, Mishra SI, Adams A, Dickerson VM, Murphy LS, Reinsch S. Reliability of spinal palpation for diagnosis of back and neck pain: a systematic review of the literature. Spine 2004;29:E413-25.
29. Stochkendahl MJ, Christensen HW, Hartvigsen J, Vach W, Haas M, Hestbæk L, Adams A, Bronfort G. Manual examination of the spine: a systematic critical literature review of reproducibility. J Manipulative Physiol Ther 2006;29:475-85.
30. Van Trijffel E, Anderegg Q, Bossuyt PM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. Man Ther 2005;10:256-69.
31. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. Clin Chem 2004;50:473-6.
32. Jones MA, Jensen G, Edwards I. Clinical reasoning in physiotherapy. In: Higgs J, Jones MA, Loftus S, Christensen N (editors). Clinical reasoning in the health professions (3rd edition). Edinburgh, UK: Elsevier/Butterworth Heinemann, 2008.
33. Rothstein JM, Echternach JL, Riddle DL. The hypothesis-oriented algorithm for clinicians II (HOAC II): a guide for patient management. Phys Ther 2003;83:455-70.
34. Rivett DA, Higgs J. Hypothesis generation in the clinical reasoning behaviour of manual therapists. J Phys Ther Educ 1997;11:40-5.
35. Van Trijffel E, Plochg T, Van Hartingsveld F, Lucas C, Oostendorp RAB. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual physical therapy: a qualitative interview study. J Man Manip Ther 2010;18:111-8.
36. Carlesso LC, Macdermid JC, Santaguida PL, Thabane L, Giulekas K, Larocque L, Millard J, Williams C, Miller J, Chesworth BM. Beliefs and practice patterns in spinal manipulation

and spinal motion palpation reported by Canadian manipulative therapists. Physiother Canada 2013;65:167-75.

37. Riddle DL. Classification and low back pain: A review of the literature and critical analysis of selected systems. Phys Ther 1996;78:708-37.
38. Fairbank J, Gwilym SE, France JC, Daffner SD, Dettori J, Hermsmeyer J, Andersson G. The role of classification of chronic low back pain. Spine 2011;36:S19-42.
39. Arab AM, Abdollahi I, Joghataei MT, Golafshani Z, Kazemnejad A. Inter- and intra-examiner reliability of single and composites of selected motion palpation and pain provocation tests for sacroiliac joint. Man Ther 2009;14:213-21.
40. Kokmeyer DJ, Van der Wurff P, Aufdemkampe G, Fickenscher TC. The reliability of multitest regimens with sacroiliac pain provocation tests. J Manipulative Physiol Ther 2002;25:42-8.
41. Robinson HS, Brox JI, Robinson R, Bjelland E, Solem S, Telje T. The reliability of selected motion- and pain provocation tests for the sacroiliac joint. Man Ther 2007;12:72-9.
42. Fritz JM, Brennan GP, Clifford SN, Hunter SJ, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. Spine 2006;31:77-82.
43. Stanton TR, Fritz JM, Hancock MJ, Latimer J, Maher CG, Wand BM, Parent EC. Evaluation of a treatment-based classification algorithm for low back pain: A cross-sectional study. Phys Ther 2011;91:496-509.
44. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, Delitto A. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. Ann Intern Med 2004;141:920-8.
45. Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, Butler D, Garber M, Allison S. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27:2835-43.
46. Fritz JM, Whitman JM, Flynn TW, Wainner RS, Childs JD. Factors related to the inability of individuals with low back pain to improve with a spinal manipulation. Phys Ther 2004;84:173-90.
47. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Arch Phys Med Rehabil 2005;89:1753-62.

48. Cook C. Key issues for manual therapy clinical practice and research in North America. Man Ther 2013;18:269-70.
49. Fritz JM, Brennan GP. Preliminary examination of a proposed treatment-based classification system for patients receiving physical therapy interventions for neck pain. Phys Ther 2007;87:513-24.
50. Cleland JA, Childs JD, Fritz JM, Whitman JM, Eberhart SL. Development of a clinical prediction rule for guiding treatment of a subgroup of patients with neck pain: Use of thoracic spine manipulation, exercise, and patient education. Phys Ther 2007;87:9-23.
51. Cleland JA, Fritz JM, Whitman JM, Heath R. Predictors of short-term outcome in people with a clinical diagnosis of cervical radiculopathy. Phys Ther 2007;87:1619-32.
52. Puentedura EJ, Cleland JA, Landers MR, Mintken PE, Louw A, Fernándes-de-las-Peñas C. J Orthop Sports Phys Ther 2012;42:577-92.
53. Schellingerhout JM, Verhagen AP, Heymans MW, Pool JJ, Vonk F, Koes B, De Vet HCW. Which subgroups of patients with non-specific neck pain are more likely to benefit from spinal manipulation, physiotherapy, or usual care? Pain 2008;139:670-80.
54. Thiel HW, Bolton JE. Predictors for immediate and global responses to chiropractic manipulation of the cervical spine. J Manipulative Physiol Ther 2008;31:172-83.
55. Tseng Y-L, Wang WTJ, Chen W-Y, Hou T-J, Chen T-C, Lieu F-K. Predictors for the immediate responders to cervical manipulation in patients with neck pain. Man Ther 2006;11:306-15.
56. Gemmell H, Miller P. Interexaminer reliability of multidimensional examination regimens for detecting spinal manipulable lesions: a systematic review. Clin Chiropr 2005;8:199-204.
57. De Hertogh WJ, Vaes PH, Vijverman V, De Cordt A, Duquet W. The clinical examination of neck pain patients: the validity of a group of tests. Man Ther 2007;12:50-5.
58. Guzman J, Hurwitz EL, Carroll LJ, Haldeman S, Côté P, Carragee EJ, Peloso PM, Van der Velde G, Holm LW, Hogg-Johnson S, Nordin M, Cassidy JD. A new conceptual model of neck pain: linking onset, course, and care. The Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. Spine 2008;33:S14-23.
59. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffet J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006;15:S192-300.

60. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain. Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res 2011;63:S240-52.
61. Schoppink LE, Van Tulder MW, Koes BW, Beurskens SA, De Bie RA. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. Phys Ther 1996;76:268-75.
62. Schellingerhout JM, Heymans MW, Verhagen AP, De Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: A systematic review. BMC Med Res Methodol 2011;11:87.
63. Greenhalgh S, Selfe J. Red flags. A guide to identifying serious pathology of the spine. Amsterdam/New York: Elsevier/Churchill Livingstone, 2006.
64. Staal JB, Hendriks EJM, Heijmans M, Kiers H, Lutgers-Boomsma AM, Rutten G, Van Tulder MW, Den Boer J, Ostelo R, Custers JWH. KNGF Guideline Low-back Pain, 2013. Available at: http://www.fysionet-evidencebased.nl/index.php/component/kngf/richtlijnen/lage-rugpijn-2013. [In Dutch] (accessed 26 July, 2013)
65. Van der El A. Orthopaedic manual therapy diagnosis. Spine and temperomandibular joints. London, UK: Jones and Bartlett Publishers, 2010.
66. Macfadyen N, Maher CG, Adams R. Number of sampling movements and manual stiffness judgements. J Manipulative Physiol Ther 1998;21:604-10.
67. International Classification of Functioning, Disability and Health: ICF. Geneva, Switzerland: World Health Organization, 2001. Available at: http://www.who.int/classifications/icf/en/. (accessed 26 July, 2013)
68. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37-46.
69. Cicchetti DV, Feinstein AR. High agreement but low Kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551-8.
70. Byrt T, Bishop J, Carlin JB. Bias, prevalence and Kappa. J Clin Epidemiol 1993;46:423-9.
71. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.

72. MacKinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. Comput Biol Med 2000;30:127-34.
73. Landis JR, Koch DG. The measurement of observer agreement for categorical data. Biometrica 1977;33:159–64.
74. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, and the QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol 2013;66:1093-1104.
75. Foster NE, Hill JC, Hay EM. Subgrouping patients with low back pain in primary care: Are we getting any better at it? Man Ther 2011;16:3-8.
76. Borkan JM, Cherkin DC. An agenda for primary care research on low back pain. Spine 1996;21:2880-4.
77. Borkan JM, Koes B, Reis S, Cherkin DC. A report from the Second International Forum for Primary Care Research on Low Back Pain. Examining priorities. Spine 1998;23:1992-6.
78. Bouter LM, Van Tulder MW, Koes BW. Methodologic issues in low back pain research in primary care. Spine 1998;23:2014-20.
79. Clinical Research Agenda for Physical Therapy. Phys Ther 2000;80:499-513.
80. Foster NE, Dziedzic KS, Van der Windt DAWM, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problem: Nationally and internationally agreed recommendations. BMC Musculoskelet Disord 2009;10:3.
81. Goldstein MS, Scalzitti DA, Craik RL, Dunn SL, Irion JM, Irrgang J, Kolobe THA, McDonough CM, Shields RK. The Revised Research Agenda for Physical Therapy. Phys Ther 2011;91:165-74.
82. Kent P, Keating JL, Leboeuf-Yde C. Research methods for subgrouping low back pain. BMC Med Res Methodol 2010;10:62.
83. Kent P, Keating J. Do primary care clinicians think that nonspecific low back pain is one condition? Spine 2004;29:1022-31.
84. Kent P, Keating J. Classification in nonspecific low back pain: What methods do primary care clinicians currently use? Spine 2005;30:1433-40.
85. Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: A systematic review. Phys Ther 2009;89:114-24.

86. May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: A systematic review. J Man Manip Ther 2009;17:36-45.
87. Kent P, Mjøsund HL, Petersen DHD. Does targeting manual therapy and/or exercise improve patient outcomes in nonspecific low back pain. A systematic review. BMC Med 2010;8:22.
88. Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Phys Ther 2010;90:843-54.
89. Patel S, Friede T, Froud R, Evans DW, Underwood M. Systematic review of randomized controlled trials of clinical prediction rules for physical therapy in low back pain. Spine 2013;38:762-9.
90. Slater SL, Ford JJ, Richards MC, Taylor NF, Surkitt LD, Hahne AJ. The effectiveness of sub-group specific manual therapy for low back pain: A systematic review. Man Ther 2012;17:201-12.
91. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med 1992;21:384-90.
92. Smith J, Bolton PS. What are the clinical criteria justifying spinal manipulative therapy for neck pain? A systematic review of randomized controlled trials. Pain Med 2013;14:460-8.
93. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6.
94. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. A systematic review. Ann Intern Med 2004;140:189-202.
95. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174:469-76.
96. Scales CD Jr, Canfield SE. Advanced topics in evidence-based urological oncology: Using results of a subgroup analysis. Urol Oncol 2011;29:462-6.
97. Cook C, Turney L, Ramirez L, Miles A, Haas S, Karakostas T. Predictive factors in poor inter-rater reliability among physical therapists. J Man Manip Ther 2002;10:200-5.

98. Riddle DL, Rothstein JM, Echternach JL. Application of the HOAC II: An episode of care for a patient with low back pain. Phys Ther 2003;83:471-85.
99. Thoomes EJ, Schmitt MS. Practical use of the HOAC II for clinical decision making and subsequent therapeutic interventions in an elite athlete with low back pain. J Orthop Sports Phys Ther 2011;41:108-17.
100. Oostendorp RAB, Rutten GM, Dommerholt J, Nijhuis-van der Sanden MW, Harting J. Guideline-based development and practice test of quality indicators for physiotherapy care in patients with neck pain. J Eval Clin Prac 2013;13:194.
101. Shirley D, Ellis E, Lee M. The response of posteroanterior lumbar stiffness to repeated loading. Man Ther 2002;7:19-25.
102. Busscher I, Van Dieën JH, Van der Veen AJ, Kingma I, Meijer GJM, Verkerke GJ, Veldhuizen AG. The effects of creep and recovery on the *in vitro* biomechanical characteristics of human multi-level thoracolumbar spinal segments. Clin Biomech 2011;26:438-44.
103. Battlé MC, Cherkin DC, Dunn R, Ciol MA, Wheeler KJ. Managing low back pain: Attitudes and treatment preferences of physical therapists. Phys Ther 1994;74:219-26.
104. Carey TS, Freburger JK, Holmes GM, Castel L, Darter J, Agans R, Kalsbeek W, Jackman A. A long way to go. Practice patterns and evidence in chronic low back pain care. Spine 2009;34:718-24.
105. Freburger JK, Carey TS, Holmes GM. Physical therapy for chronic low back pain in North Carolina: Overuse, underuse, or misuse? Phys Ther 2011;91:484-95.
106. Goode AP, Freburger J, Carey T. Prevalence, practice patterns, and evidence for chronic neck pain. Arthritis Care Res 2010;62:1594-1601.
107. Gracey JH, McDonough SM, Baxter GD. Physiotherapy management of low back pain. A survey of current practice in Northern Ireland. Spine 2002;27:406-11.
108. Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. Phys Ther 1997;77:145-54.
109. Li LC, Bombardier C. Physical therapy management of low back pain: An exploratory survey of therapist approaches. Phys Ther 2001;81:1018-28.
110. Liddle SD, Baxter GD, Gracey JH. Physiotherapists' use of advice and exercise for the management of chronic low back pain: A national survey. Man Ther 2009;14:189-96.

111. Mikhail C, Korner-Bitensky N, Rossignol M, Dumas J-P. Physical therapists' use of interventions with high evidence of effectiveness in the management of a hypothetical typical patient with acute low back pain. Phys Ther 2005;85:1151-67.

112. Mielenz TJ, Carey TS, Dyrek DA, Harris BA, Garrett JM, Darter JD. Physical therapy utilization by patients with acute low back pain. Phys Ther 1997;77:1040-51.

113. Van Baar ME, Dekker J, Bosveld W. A survey of physical therapy goals and interventions for patients with back and knee pain. Phys Ther 1998;78:33-42.

114. Adams G, Sim J. A survey of UK manual therapists' practice of and attitudes towards manipulation and its complications. Physiother Res Int 1998;3:206-27.

115. Jull G. Use of high and low velocity cervical manipulative therapy procedures by Australian manipulative physiotherapists. Aust J Physiother 2002;48:189-93.

116. Hurley L, Yardley K, Gross AR, Hendry L, McLaughlin L. A survey to examine attitudes and patterns of practice of physiotherapists who perform cervical spinal manipulation. Man Ther 2002;7:10-8.

117. Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257-68.

118. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. J Clin Epidemiol 2000;53:499-503.

119. Zusman M. There's something about passive movement…. Med Hypotheses 2010;75:106-10.

120. Hurwitz EL, Morgenstern H, Harber P, Kominski GF, Yu F, Adams AH. A randomized trial of chiropractic manipulation and mobilization for patients with neck pain: Clinical outcomes from the UCLA neck-pain study. Am J Public Health 2002;92:1634-41.

121. Leaver AM, Maher CG, Herbert RD, Latimer J, McAuley JH, Jull G, Refshauge KM. A randomized controlled trial comparing manipulation with mobilization for recent onset neck pain. Arch Phys Med Rehabil 2010;91:1313-8.

122. Dunning JR, Cleland JA, Waldrop MA, Arnot C, Young I, Turner M, Sigurdsson G. Upper cervical and upper thoracic thrust manipulation versus mobilization in patients with mechanical neck pain: A multicenter randomized clinical trial. J Orthop Sports Phys Ther 2012;42:5-18.

123. Cook C, Learman K, Showalter C, Kabbaz V, O'Halloran B. Early use of thrust manipulation versus non-thrust manipulation: A randomized clinical trial. Man Ther 2013;18:191-8.

Appendix

Formulas

			Examiner 1		
			Indication		
			Positive	Negative	
Examiner 2	**Indication**	Positive	*a*	*b*	g_1
		Negative	*c*	*d*	g_2
			f_1	f_2	*n*

$$p_o = \frac{a + d}{n}$$

$$p_e = \frac{\left(\frac{f_1 \times g_1}{n}\right) + \left(\frac{f_2 \times g_2}{n}\right)}{n}$$

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$$\mathrm{SE}\ (\kappa) = \sqrt{\frac{p_o(1 - p_o)}{n(1 - p_e)^2}}$$

$$95\%\ \mathrm{CI}\ (\kappa) = \kappa - 1.96 \times \mathrm{SE}\ (\kappa) \text{ to } \kappa + 1.96 \times \mathrm{SE}\ (\kappa)$$

$$\mathrm{BI} = \frac{|b - c|}{n}$$

$$\mathrm{PI} = \frac{|a - d|}{n}$$

$$\mathrm{p_{neg}} = \frac{d}{\left(\frac{f_2 + g_2}{2}\right)}$$

$$\mathrm{p_{pos}} = \frac{a}{\left(\frac{f_1 + g_1}{2}\right)}$$

BI: bias index, CI: confidence interval, κ: kappa, p_e: expected agreement by chance, PI: prevalence index, p_{neg}: proportion of agreement on negative indications, p_o: observed agreement, p_{pos}: proportion of agreement on positive indications, SE: standard error

Chapter 7

Ultrasonography and magnetic resonance imaging were not appropriate for measuring *in vivo* time-dependent changes in synovial fluid volume after passive joint movements. A short communication

Emiel van Trijffel, Michel de Maeseneer, Luca Buzzatti, Aldo Scafoglieri, Erik Cattrysse

Abstract

Background: The mechanisms behind passive movement-based joint interventions commonly used in manual therapy such as mobilisation and manipulation are largely unknown. Biomechanical mechanisms have only scarcely been investigated and data on the role of specific structures responsible for *in vivo* time-dependent changes in biomechanical behaviour of human joints of either the spine or the extremities after passive joint movements are not available. We hypothesised that potential changes in biomechanical properties of joints after passive movements would be due to immediate changes in the volume or distribution of synovial fluid.

Methods: We conducted a series of experiments in three healthy subjects using ultrasonography (US) and magnetic resonance imaging (MRI) for visualising and measuring time-dependent changes in synovial fluid volume in joints of the upper cervical spine, the knee joints, and the metacarpophalangeal joints up to 60 minutes before and after passive joint motion assessment, mobilisation, and high-velocity, low-amplitude thrust manipulation.

Results: MRI could not detect any fluid in the articular space of the lateral atlanto-axial joints. Using US, imaging of the palmar recess of the metacarpophalangeal joints of the second and third fingers was considered insufficiently reproducible preventing us from obtaining complete data to visualise and measure the volume and distribution of fluid in both the palmar and dorsal recesses as related to the synovial fluid in the intra-articular space. Thirty minutes after mobilisation of the knee in one subject, the antero-posterior diameter of the suprapatellar recess was decreased from 11.0 to 9.0 millimeters.

Conclusions: We conclude that current US and MRI techniques are not appropriate for visualisation and measurement of *in vivo* time-dependent changes, if any, in the volume of synovial fluid after passive movement-based joint interventions. New, innovative research is needed to generate evidence on the biomechanical effects of passive movement-based joint interventions.

Introduction

It is increasingly proposed to support treatment interventions with evidence of their mechanisms alongside evidence of effectiveness.[1;2] There is evidence that passive joint mobilisations and high-velocity, low-amplitude thrust manipulations are effective, as well as cost-effective, in patients with non-specific neck or low-back pain although no more effective than other treatment modalities.[3-8] There is also mounting evidence that these interventions are effective in patients with upper or lower extremity disorders.[9;10] However, the mechanisms behind mobilisation and manipulation remain elusive.[11;12]

Spinal joint manipulations are considered mechanical events.[13] They generate immediate, albeit small, effects on range of motion, especially in the cervical spine.[14] Better immediate improvements in pain scores were obtained from single session cervical manipulations as compared to mobilisations.[15] There is growing evidence that the effects of spinal manipulation are mediated by the central nervous system through somatosensory activation and addressing disordered sensorimotor integration and motor control.[16-18]

Evidence for a mechanical effect of passive mobilising joint movements using graded loading or stretching of connective tissues is lacking.[19] Passive accessory cervical joint mobilisation was found to activate central nervous system mechanisms responsible for pain control and autonomic function.[20] In addition, in a systematic review, it was concluded that single sessions of spinal joint mobilisations have immediate, within-session effects on pain.[21]

To summarise, evidence for the explanation of effects from passive movement-based interventions seems to be in favour of neurophysiological mechanisms.[19] A model has been proposed suggesting that a mechanical force, i.e., a joint manipulation or mobilisation, initiates a cascade of neurophysiological responses from the peripheral and central nervous system associated with pain relief.[11]

Biomechanical mechanisms have only scarcely been investigated. In addition, only very few studies have investigated the mechanical behaviour of human musculoskeletal structures as a function of time. Shirley et al[22] reported that stiffness responses to repeated mechanical postero-anterior loading of lumbar motion segments returned to the pre-testing state within five minutes. On the other hand, a 30-minute recovery period after 30 minutes of *in vitro* creep loading of the lumbar spine was not sufficient to return to the baseline situation.[23] Despite these conflicting results, detailed data on the role of specific structures responsible for *in vivo* time-dependent changes in biomechanical behaviour of human joints of either the spine or the extremities after passive joint movements are not available.

Synovial fluid of joints acts as a lubricant as well as a biochemical depot through which nutrients and cytokines traverse.[24] Intra-articular fluid pressure is directly affected by joint movements and active or passive motion can increase this pressure above atmospheric pressure creating a net flow of fluid transport out of the joint cavity.[25] Movements also influence the distribution of synovial joint fluid. For example, in *ex vivo* rabbit knees, fluid moved from the anterior to the posterior bursae during increasing flexion as detected by contrast-enhanced micro-computed tomography and X-ray imaging.[26] However, *in vivo* data on the flow, volume, or distribution of synovial fluid during or after passive joint movements in humans are lacking.

Mechanical effects of passive joint interventions such as stiffness changes could be expected to occur relatively immediately.[27] Biomechanically, synovia behaves as a viscoelastic fluid due to the presence of hyaluron molecules, but it is also known to possess viscous properties in processes that are steady.[28] It is assumed that changes in the viscosity of synovial fluid from rest to movement and vice versa can occur within seconds to several minutes.[29;30] It is, however, unknown whether passive joint movements commonly used in manual therapy are associated with changes in the behaviour of synovial fluid and how these changes, if any, develop over time.

We hypothesised that potential changes in biomechanical properties of joints after passive movements are due to immediate changes in intra-articular joint pressure and, consequently, in volume or distribution of synovial fluid. We used ultrasonography (US) and magnetic resonance imaging (MRI) for visualising and measuring *in vivo* time-dependent changes, if any, in synovial fluid volume in joints of the upper cervical spine, the knee joints, and the metacarpophalangeal joints of human subjects after passive joint motion assessment, mobilisation, and thrust manipulation.

Methods

Design

We conducted repeated measurements in human subjects using US and MRI before and after passive motion assessment, mobilisation, and high-velocity, low-amplitude thrust manipulation of joints of the upper cervical spine, the knee joints, and the metacarpophalangeal joints. The experiments took place at the Department of Experimental Anatomy of the Vrije Universiteit Brussel, Brussels, Belgium. The Medical Ethics Committee of the University Hospital of Brussels, Belgium, stated in writing that the study was exempt from approval.

Participants

Three healthy male subjects A-C, respectively 27, 46, and 25 years of age, without complaints of their neck, knees, or fingers during the last six months or any trauma, fractures, or surgery in these regions in the past volunteered to participate in the experiments. All subjects gave written informed consent prior to participation.

Procedures

All procedures started by instructing subjects to rest supine for 30 minutes. During this period, after 15 and 30 minutes, US imaging was used to visualise and measure the diameter of the synovial recesses of the knee joints and metacarpophalangeal joints while MRI was used to detect synovial fluid in the lateral atlanto-axial joints. After 30 minutes, passive joint movements were performed by a physiotherapist (EvT) with over 15 years of experience in manual therapy. Passive joint motion assessment, mobilisation, and thrust manipulation were executed in accordance with current international textbook guidelines.[31;32] Each movement during passive motion assessment was performed until the end of the range of movement was perceived by the therapist. All these movements were produced three times in each motion direction. Assessments per joint as well as in the upper cervical region lasted a maximum of five minutes. Passive mobilisations were performed as grade IV movements with end-positions maintained for 10 seconds in all motion directions while thrust manipulations were grade V techniques.[33] Total duration of the mobilisation protocols per joint, as well as in the upper cervical region, was set at 10 minutes. After the movement interventions, participants were again instructed to rest supine for 60 minutes. US or MRI were performed directly after the interventions, after 30 minutes, and after 60 minutes. For MRI (Philips Achieva 3.0 Tesla), sagittal T2-weighted (TSE, TR: 3741 ms, TE: 100 ms, 3 mm thickness) and STIR (TR: 4519 ms, TE: 80 ms, 3 mm thickness) sequences were applied while US (Aloka F75) used an 18 MHz (1.5D) matrix probe. All US and MRI were carried out and interpreted by an expert radiologist (MdM) with 20 years of experience in musculoskeletal imaging.

Upper cervical spine

In subject A, MRI was used to visualise the synovial fluid in the articular space of the left and right lateral atlanto-axial joints. Passive motion assessment included contralateral three-dimensional left and right regional flexion and extension, three-dimensional left and right segmental flexion and extension at the C2-C3 motion segment, left and right rotation at the atlanto-axial segment, and physiological flexion, extension, and left and right contralateral rotation-lateral flexion at the atlanto-occipital segment. Sixty minutes after finishing the assessment procedures, these same movements were performed as joint mobilisations. No joint manipulations were applied to the upper cervical region.

Knee joints

Optimal technique, position, and localisation for visualising and measuring the antero-posterior diameter of the suprapatellar recess of the left and right knee joints using US were first tested in subject B. We chose to focus on the suprapatellar recess in isolation and not related to other recesses or the synovial fluid in the intra-articular joint space. Subsequently, the imaging and intervention protocol was applied to subject C. During US imaging, the subject was lying supine with the knee in a supported 20 degrees of flexion position. The antero-posterior diameter was calculated as the mean from two consecutive measurement occasions of two distances constituting the proximal and distal boundaries of the central one-third of the recess (Figure).

Passive mobilisations of the left knee joint were performed as physiological flexion, extension, internal and external rotation, accessory posterior tibial glide in 20° flexion and in end-flexion and ventral tibial glide in 20° flexion and in end-extension, traction and compression in 20° flexion, and inferior glide of the patella in 20° flexion. The right knee was used as a control. Sixty minutes after finishing the assessment and mobilisation procedures, subject C was additionally asked to perform an active loading regimen consisting of three sets of three minutes walking plus 15 squats, with a total duration of 12 minutes. Immediately after, US imaging was conducted to the right knee. No thrust manipulations were applied to the knee joints.

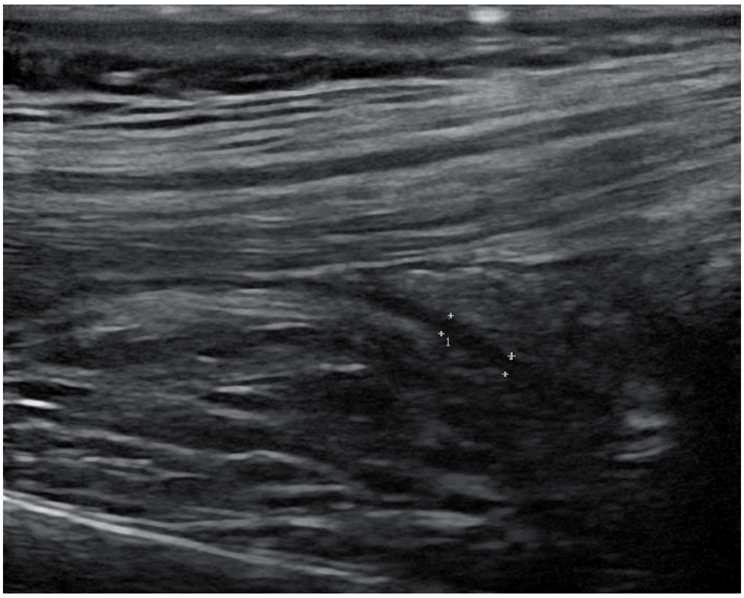

Figure. Definition of two distances constituting the proximal and distal boundaries of the central one-third of the suprapatellar recess for measuring its antero-posterior diameter

Metacarpophalangeal joints

We first tested whether US would allow visualisation of a known change in the volume of the intra-articular fluid. Two cc of a physiological saline solution was injected in the third metacarpophalangeal joint of a well-preserved, white Caucasian, adult embalmed cadaver. A clear increase in the volume of the dorsal recess of the joint was visible, leading us to pursue our experiments in an *in vivo* environment.

Optimal technique, position, and localisation for visualising and measuring the dorsal-palmar diameter of the palmar and dorsal recesses of the metacarpophalangeal joints of the second and third fingers of both hands using US were again first tested in subject B. At this stage, we aimed to measure changes, if any, in fluid volume in both the palmar and dorsal recesses as related to each other as well as to the synovial fluid in the intra-articular space to allow for examination of changes in fluid distribution. The imaging and intervention protocol was subsequently applied to subject C. During US imaging, the subject was seated on a chair with the forearm and hand resting on the examination table and the elbow in 90 degrees of flexion. Imaging of the palmar recess took place with the forearm in a supinated position while a pronated position was used for the dorsal recess.

Passive motion assessment of the metacarpophalangeal joint of the second finger of the left hand consisted of physiological flexion and extension, accessory palmar phalangeal glide in neutral position and in end-flexion and dorsal phalangeal glide in neutral position and in end-extension, and traction and compression in neutral position. These movements were then applied as mobilisations of the metacarpophalangeal joint of the third finger of this hand. The metacarpophalangeal joint of the second finger of the right hand was used for single thrust manipulations in traction and flexion directions while the third finger of this hand was used as a control.

In an attempt to directly measure the volume of the intra-articular fluid in the metacarpophalangeal joint, we performed a further experiment using MRI and NeuroScape™ software (Olea Medical, La Ciotat, France) in a series of images obtained from a randomly selected, anonymised patient by courtesy of the Department of Radiology of the University Hospital of Brussels, Belgium. Imaging and interpretation were carried out by a musculoskeletal radiologist (MdM) assisted by a neuroradiologist.

Results

Upper cervical spine

Using MRI in subject A, it was not possible to demonstrate any fluid in the articular space of either the left or right lateral atlanto-axial joint.

Knee joints

The results of measuring the antero-posterior diameter of the suprapatellar recess of the left knee joint in subject B using US before and after the mobilisation interventions, with the right knee as a control, are presented in the Table.

Table. Antero-posterior diameter of the suprapatellar recess of the left knee joint in subject B using ultrasonography before and after the mobilisation interventions, with the right knee as a control

Joint	15' pre-intervention	0' pre-intervention	0' post-intervention	30' post-intervention	60' post-intervention
Left knee (mobilisation)	11.0	10.0	11.0	9.0	10.0
Right knee (control)	11.5	9.5	10.0	10.0	10.5

The mean antero-posterior diameter of the recesses of both knee joints was decreased after the initial 30 minute resting period. Thirty minutes after the mobilisation interventions to the left knee, the diameter was further decreased from 11.0 to 9.0 millimeters. The mean antero-posterior diameter of the recess of the right knee immediately after the active loading regimen was 11.0 millimeters.

Metacarpohalangeal joints

Using US in subject C, imaging of the palmar recess of the metacarpophalangeal joints of the second and third fingers of both hands was considered insufficiently reproducible. Consequently, we were not able to obtain measurements of fluid volume in both the palmar and dorsal recesses as related to each other as well as to the synovial fluid in the intra-articular space. No additional fluid was visible at the dorsal recess after the passive joint interventions.

Finally, it was not possible to directly measure the volume of the intra-articular fluid in the metacarpophalangeal joint using MRI and NeuroScape™ software. It became apparent that the software was not able to delineate the intra-articular fluid and distinguish it from other surrounding fluid-containing tissues such as blood vessels.

Discussion

In this series of basic science experiments, US and MRI were used for visualising and measuring *in vivo* time-dependent changes in synovial fluid volume in human joints of the upper cervical spine, the knee joints, and the metacarpophalangeal joints as potential, immediately occurring biomechanical effects of passive joint movements commonly used in manual therapy. Using MRI, we were not able to visualise any fluid in the articular space of the lateral atlanto-axial joints while US was not sufficiently reproducible to allow a complete analysis of the volume and distribution of fluid in the recesses of the metacarpophalangeal joints as related to the synovial fluid in the intra-articular space. The mean antero-posterior diameter of the suprapatellar recess of the knee decreased during the initial 30 minute resting period and it was further decreased 30 minutes after the mobilisation interventions as measured using US. However, it must be noted that no estimates of measurement error were determined.

We used MRI to visualise fluid in the articular space of the joints of the atlanto-axial motion segment which possess a large, loose capsule and a large rotational mobility. Effects of mobilisation interventions on synovial fluid may, therefore, be easier to detect. MRI has been used successfully in human subjects to quantify cavitation and gapping of lumbar zygapophyseal joints during thrust manipulation by determining the difference between pre- and post-intervention joint space measurements.[34;35] However, synovial fluid in the intra-articular space was not the focus in these investigations. Our preliminary results indicate that MRI is not suitable for detecting fluid in the intra-articular space of intervertebral joints.

Our results for the knee joints showed a decrease in the diameter of the suprapatellar recess after the initial resting period as measured using US. This was contrary to what was expected based on theoretical grounds as lowering of the intra-articular fluid pressure would be associated with a net flow of fluid transport into the joint cavity.[25] The further decrease of the diameter of this recess after the mobilisation interventions was, on the other hand, a confirmation of the theory as it could be related to an increase of intra-articular pressure and a net flow of fluid out of the cavity. However, these findings should be interpreted with caution as one has to realise that the diameter of a recess is not a direct representation of the volume of the fluid contained because of its two-dimensional representation. In addition, redistribution of fluid may be an even more important influencing factor when exploring changes in volume of synovial fluid across recesses and intra-articular space, especially in the knee with its many recesses and bursae.

In our next experiment, we attempted to measure the diameter of both the palmar and dorsal recesses of the metacarpophalangeal joints as related to each other as well as to the fluid in the intra-articular space. These data could have provided evidence on the role of redistribution of fluid after passive movements. Unfortunately, measurements of the palmar recess were not reproducible. Replication of these experiments in a larger sample of subjects could perhaps provide more reproducible and useful data.

We were not able to support our hypothesis that changes in biomechanical properties of joints after passive movements are related to immediate changes in volume or distribution of synovial fluid. These changes may indeed not occur at all. Alternatively, they may be too small to detect with current imaging techniques. We observed that using US it was possible to detect a change in fluid volume of two cc in the dorsal recess of a metacarpophalangeal joint, but changes, if any, induced by passive movements may be far smaller.

There is limited evidence that passive movements do induce changes in properties of synovial fluid. For instance, cyclic variation in intra-articular pressure, changes in synovial fluid volume, and increased trans-synovial transport have been observed during continuous passive motion regimens in the rabbit[36;37] and human[38] knee. Recently, lower cervical thrust manipulation and thoracic manipulation produced mechanical strains that were innocuous to joint and surrounding tissues in healthy human subjects.[39] In short, biomechanical mechanisms of passive movement-based interventions in manual therapy are still a largely unexplored research area.

While current imaging techniques may not enable us to pursue with the hypothesis of changes in biomechanical properties related to synovial fluid, focusing on kinematical parameters such as the quantification of three-dimensional joint motion behaviour after passive movements may provide a new and promising research direction.[40] Alternatively, US elastography as a non-invasive, low cost, and real-time access diagnostic technique for disorders of tendons, ligaments, and muscles may expand in future towards the qualitative and quantitative three-dimensional evaluation of biomechanical properties of capsulo-ligamentous joint structures.[41]

Evidence on biomechanical mechanisms of passive movements could inform researchers and clinicians in two areas. First, it is currently uncertain whether mobilisation or manipulation, or both, should be preferred in non-specific spinal pain. Results of randomised controlled trials are undecided so far.[42-49] Knowledge of the mechanisms through which mobilisation and manipulation each generate their effects, either biomechanical or neurophysiological, could guide the targeting of treatment to patients in clinical trials and clinical practice by connecting clinical findings, treatment intervention, and subsequent patient-oriented outcomes.[50]

Second, researchers in the field of reliability of passive joint motion assessment have been confronted with the difficult methodological issue of ensuring stability of the characteristic under study during the research, i.e., between sessions of assessments.[51-53] Instability of the joint's mobility as a result of natural variation over time or as an effect of the assessment procedure itself could produce underestimated outcomes of reliability. Knowledge of the time-dependent biomechanical properties and behaviour of joints following passive joint motion assessment will help researchers to define the number of tests, the number of movement repetitions, forces applied in the end-position, motion directions, and time intervals in order to achieve an unbiased estimation of reliability.

Limitations of this study

These pilot experiments had several methodological limitations. First, we chose to include only healthy subjects because data on the time-dependent biomechanical behaviour under normal, physiological conditions, which could serve as a reference when pursuing the investigation further in patients, were not available. Including patients more representative of clinical practice in manual therapy, e.g., those with (non-specific) musculoskeletal pain, or patients with osteoarthritis or rheumatoid arthritis in which the synovial tissue and synovial fluid is primarily affected[54-55], could have led to different conclusions. Second, we only included a limited number of subjects. A larger number of subjects could have provided richer data and other conclusions. Third, judgements of US and MRI images are inevitably susceptible to measurement error. Although no evidence exists on the intra-examiner reliability or test-retest reliability of judgements of these images as related to (synovial) fluid, at this stage we did not estimate any measurement error. In particular, our measurement data on the diameter of the suprapatellar recess of the knee could have been influenced by measurement error.

Conclusions and recommendations

We conclude that current techniques for US and MRI are not appropriate for visualisation and measurement of *in vivo* time-dependent changes, if any, in the volume of synovial fluid after passive movement-based joint interventions. Alternative lines of innovative research need to be explored to generate evidence on the biomechanical effects of passive joint interventions commonly used in manual therapy.

References

1. Howick J, Glasziou P, Aronson JK. Evidence-based mechanistic reasoning. J R Soc Med 2010;103:433.
2. Clarke B, Gillies D, Illari P, Russo F, Williamson J. The evidence that evidence-based medicine omits. Prev Med 2013;57:745-7.
3. Bronfort G, Haas M, Evans R, Leininger B, Triano J. Effectiveness of manual therapies: The UK evidence report. Chiropr Osteopat 2010;18:3.
4. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Brønfort G, Hoving JL. Manipulation or mobilisation for neck pain. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD004249. DOI:10.1002/14651858.CD004249.pub3.
5. Rubinstein SM, Van Middelkoop M, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for chronic low-back pain. Cochrane Database of Systematic Reviews 2011, Issue 2. Art. No.: CD008112.
6. Rubinstein SM, Terwee CB, Assendelft WJJ, De Boer MR, Van Tulder MW. Spinal manipulative therapy for acute low-back pain. Cochrane Database of Systematic Reviews 2012, Issue 9. Art. No.: CD008880.
7. Michaleff ZA, Lin C-WC, Maher CG, Van Tulder MW. Spinal manipulation epidemiology: Systematic review of cost-effectiveness studies. J Electromyogr Kinesiol 2012;22:655-62.
8. Tserstvadze A, Clar C, Court R, Clarke A, Mistry H, Sutcliffe P. Cost-effectiveness of manual therapy for the management of musculoskeletal conditions: A systematic review and narrative synthesis of evidence from randomized controlled trials. J Manipulative Physiol Ther 2014;37:343-62.
9. Brantingham JW, Bonnefin D, Perle SM, Cassa TK, Globe G, Pribicevic M, Hicks M, Korporaal C. Manipulative therapy for lower extremity conditions: Update of a literature review. J Manipulative Physiol Ther 2012;35:127-66.
10. Brantingham JW, Cassa TK, Bonnefin D, Pribicevic M, Robb A, Pollard H, Tong V, Korporaal C. Manipulative and multimodal therapy for upper extremity and temporomandibular disorders: A systematic review. J Manipulative Physiol Ther 2013;36:143-201.

11. Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: A comprehensive model. Man Ther 2009;14:531-8.
12. Zusman M. There's something about passive movement…. Med Hypotheses 2010;75:106-10.
13. Herzog W. The biomechanics of spinal manipulation. J Bodyw Mov Ther 2010;14:280-6.
14. Millan M, Leboeuf-Yde C, Budgell B, Descarreaux M, Amorim M-A. The effect of spinal manipulative therapy on spinal range of motion: A systematic literature review. Chirop Man Therap 2012;20:23.
15. Vernon H, Humphreys BK. Chronic mechanical neck pain in adults treated by manual therapy: A systematic review of change scores in randomized controlled trials of a single session. J Man Manip Ther 2008;16:E42-52.
16. Haavik H, Murphy B. The role of spinal manipulation in addressing disordered sensorimotor integration and altered motor control. J Electromyogr Kinesiol 2012;22:768-76.
17. Pickar JG, Bolton PS. Spinal manipulative therapy and somatosensory activation. J Electromyogr Kinesiol 2012;22:785-94.
18. Savva C, Giakas G, Efstathiou M. The role of the descending inhibitory pain mechanism in musculoskeletal pain following high-velocity, low-amplitude thrust manipulation. A review of the literature. J Back Musculoskelet Rehabil 2014;27:377-82.
19. Zusman M. A note to the musculoskeletal physiotherapist. J Back Musculoskelet Rehabil 2012;25;103-107.
20. Schmid A, Brunner F, Wright A, Bachmann LM. Paradigm shift in manual therapy? Evidence for a central nervous system component in the response to passive cervical joint mobilisation. Man Ther 2008;13:387-96.
21. Slaven EJ, Goode AP, Coronado RA, Poole C, Hegedus EJ. The relative effectiveness of segment specific level and non-specific level spinal joint mobilization on pain and range of motion: Results of a systematic review and meta-analysis. J Man Manip Ther 2013;21:7-17.
22. Shirley D, Ellis E, Lee M. The response of posteroanterior lumbar stiffness to repeated loading. Man Ther 2002;7:19-25.

23. Busscher I, Van Dieën JH, Van der Veen AJ, Kingma I, Meijer GJM, Verkerke GJ, Veldhuizen AG. The effects of creep and recovery on the *in vitro* biomechanical characteristics of human multi-level thoracolumbar spinal segments. Clin Biomech 2011;26:438-44.
24. Tamer TM. Hyaluronan and synovial joint: Function, distribution and healing. Interdiscip Toxicol 2013;6:111-25.
25. Levick JR, McDonald JN. Fluid movement across synovium in healthy joints: Role of synovial fluid macromolecules. Ann Rheum Dis 1995;54:417-23.
26. McCarty WJ, Masuda K, Sah RL. Fluid movement and joint capsule strains due to flexion in rabbit knees. J Biomech 2011;44:2761-7.
27. Snodgrass SJ, Haskins R, Rivett DA. A structured review of spinal stiffness as a kinesiological outcome of manipulation: Its measurement and utility in diagnosis, prognosis and treatment decision-making. J Electromyogr Kinesiol 2012;22:708-23.
28. Hron J, Málek J, Pustèjovská P, Rajagopal KR. On the modeling of the synovial fluid. Adv Tribol 2010;2010.
29. Palfrey AJ, Davies DV. Immediate viscosity of synovial fluid. J Appl Physiol 1968;25:672-8.
30. Petrtyl M, Lisal J, Danesova J. Biomechanical properties of synovial fluid in/between peripheral zones of articular cartilage. In: Pignatello R (editor). Biomaterials – Physics and chemistry 2011:11.
31. Van der El A. Orthopaedic manual therapy diagnosis. Spine and temperomandibular joints. London, UK: Jones and Bartlett Publishers, 2010.
32. Kaltenborn FM. Manual mobilization of the joints. Volume I: The extremities (8th edition). Oslo, Norway: Olaf Norlis Bokhandel, 2014.
33. Hengeveld E, Banks K (editors). Maitland's peripheral manipulation. Management of neuromusculoskeletal disorders – Volume two (5th edition). Edinburgh: Churchill Livingstone, 2014.
34. Cramer GD, Ross K, Raju PK, Cambron J, Cantu JA, Bora P, Dexheimer J, McKinnis R, Habeck AR, Selby S, Pocius JD, Gregerson D. Quantification of cavitation and gapping of lumbar zygapophyseal joints during spinal manipulative therapy. J Manipulative Physiol Ther 2012;35:614-21.
35. Cramer GD, Cambron J, Cantu JA, Dexheimer JM, Pocius JD, Gregerson D, Fergus M, McKinnis R, Grieve TJ. Magnetic resonance imaging zygapophyseal joint space changes

(gapping) in low back pain patients following spinal manipulation and side posture positioning: A randomized controlled mechanisms trial with blinding. J Manipulative Physiol Ther 2013;36:203-17.

36. O'Driscoll SW, Kumar A, Salter RB. The effect of continuous passive motion on the clearance of a hemarthrosis from a synovial joint. An experimental investigation in the rabbit. Clin Orthop Relat Res 1983;176:305-11.
37. Danzig LA, Hargens AR, Gershuni DH, Skyhar MJ, Sfakianos PN, Akeson WH. Increased transsynovial transport with continuous passive motion. J Orthop Res 1987;5:409-13.
38. Pedowitz RA, Gershuni DH, Crenshaw AG, Petras SL, Danzig LA, Hargens AR. Intraarticular pressure during continuous passive motion of the human knee. J Orthop Res 1989;7:530-7.
39. Achalandobaso A, Plaza-Manzano G, Lomas-Vega R, Martínez-Amat A, Camacho MV, Gassó M, Hita-Contreras F, Molina F. Tissue damage markers after a spinal manipulation in healthy subjects: A preliminary report of a randomized controlled trial. Dis Markers 2014;2014:815379.
40. Cescon C, Cattrysse E, Barbero M. Methodological analysis of finite helical axis behavior in cervical kinematics. J Electromyogr Kinesiol 2014;24:628-35.
41. Klauser AS, Miyamoto H, Bellmann-Weiler R, Feuchtner GM, Wick MC, Jaschke WR. Sonoelastography: Musculoskeletal applications. Radiology 2014;272:622-33.
42. Hurwitz EL, Morgenstern H, Harber P, Kominski GF, Yu F, Adams AH. A randomized trial of chiropractic manipulation and mobilization for patients with neck pain: Clinical outcomes from the UCLA neck-pain study. Am J Public Health 2002;92:1634-41.
43. Leaver AM, Maher CG, Herbert RD, Latimer J, McAuley JH, Jull G, Refshauge KM. A randomized controlled trial comparing manipulation with mobilization for recent onset neck pain. Arch Phys Med Rehabil 2010;91:1313-8.
44. Dunning JR, Cleland JA, Waldrop MA, Arnot C, Young I, Turner M, Sigurdsson G. Upper cervical and upper thoracic thrust manipulation versus mobilization in patients with mechanical neck pain: A multicenter randomized clinical trial. J Orthop Sports Phys Ther 2012;42:5-18.
45. Cook C, Learman K, Showalter C, Kabbaz V, O'Halloran B. Early use of thrust manipulation versus non-thrust manipulation: A randomized clinical trial. Man Ther 2013;18:191-8.

46. Izquierdo Pérez H, Alonso Perez JL, Gil Martinez A, La Touche R, Lerma-Lara S, Commeaux Gonzalez N, Arribas Perez H, Bishop MD, Fernández-Carnero J. Is one better than another?: A randomized clinical trial of manual therapy for patients with chronic neck pain. Man Ther 2014;19:215-21.
47. Learman K, Showalter C, O'Halloran B, Donaldson M, Cook C. No differences in outcome in people with low back pain who met the clinical prediction rule for lumbar spine manipulation when a pragmatic non-thrust manipulation was used as a comparator. Physiother Can 2014;66:359-66.
48. Salom-Moreno J, Ortega-Santiago R, Cleland JA, Palacios-Ceña M, Truyols-Domínguez S, Fernández-de-las-Peñas C. Immediate changes in neck pain intensity and widespread pressure pain sensitivity in patients with bilateral chronic mechanical neck pain: A randomized controlled trial of thoracic thrust manipulation vs non-thrust mobilization. J Manipulative Physiol Ther 2014;37:312-19.
49. Lopez-Lopez A, Alonso Perez JL, González Gutierez JL, La Touche R, Lerma Lara S, Izquierdo H, Fernández-Carnero J. Mobilization versus manipulations versus sustain apophyseal natural glide techniques and interaction with psychological factors for patients with chronic neck pain: Randomized controlled trial. Eur J Phys Rehabil Med 2015;51:121-32.
50. Van Trijffel E, Lindeboom R, Bossuyt PMM, Schmitt MA, Lucas C, Koes BW, Oostendorp R. Indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain: Protocol of an inter-examiner reliability study among manual therapists. Chiropr Man Therap 2014;22:22.
51. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69.
52. Van de Pol RJ, Van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: A systematic review. J Physiother 2010;56:7-17.
53. Van Trijffel E, Van de Pol RJ, Oostendorp RAB, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review. J Physiother 2010;56:223-35.

54. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. New Engl J Med 2011;365:2205-19.
55. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol 2010;6:625-35.

Chapter 8

Summary

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

Summary

From a scientific perspective, uncertainty exists about the use and the value of passive intervertebral motion (PIVM) assessment of the spine within clinical diagnostics and decision-making in manual therapy. Against this background, **Chapter 1** describes the two main objectives of the research reported in this thesis: (1) to evaluate the inter-examiner reliability of passive joint motion assessment of the spine and the extremities and (2) to examine the role and position of PIVM within the process of clinical reasoning and decision-making in clinical practice in manual therapy in patients with spine-related disorders.

Chapter 2 describes a systematic review of the inter-examiner reliability of segmental PIVM assessment of the cervical (motion segments C0-T4) and lumbar (T12-S1) spine. Nineteen studies published up to March 31, 2004, were included of which nine described the results for motion assessment of the cervical spine and 10 described those of the lumbar spine. Inter-examiner reliability ranged from Cohen's kappa -0.32, for seated chiropractic assessment of lateral flexion and rotation mobility of motion segment T3-T4 in patients with non-musculoskeletal conditions, to weighted kappa 0.75, for physiotherapists using side-lying examination of motion segment L5-S1 in angular and translational movement directions in low-back pain patients. Overall, reliability was poor to fair (kappa <0.00 to 0.40). Three studies had a low risk of bias of which one found fair to moderate reliability (kappa 0.28 to 0.43) for judgements of stiffness of motion segments C1-C2, C2-C3, C7-T1, and of the first rib. Four studies used representative patients as study participants. Assessment of motion segments C1-C2 and C2-C3 consistently reached at least fair reliability (kappa >0.21). Judgements of the quantity and the quality of motion were equally (un)reliable. We concluded that the inter-examiner reliability of segmental PIVM assessment of the cervical and lumbar spine is poor as is the methodological quality of the included studies.

Chapter 3 presents two systematic reviews of the inter-examiner reliability of passive movement assessment of joints of the upper and lower extremity. We used methods very similar to those described in Chapter 2. **Chapter 3a** concerns the reliability of movement assessment of the shoulder, elbow, and wrist-hand-fingers based on studies published up to July 1, 2009. Twenty-one studies were included of which 13 investigated the shoulder, two investigated the elbow, four investigated wrist movements, one investigated phalangeal joint movements, and one investigated thumb movements. Eleven studies demonstrated acceptable reliability (ICC >0.75). Reliability varied considerably with the method of measurement and ICC ranged from 0.26 (95% CI -0.01 to 0.69), for measuring the physiological range of internal shoulder rotation using vision in shoulder patients, to 0.99 (95% CI 0.98 to 1.0), for measurements of passive torque-controlled physiological range of finger and thumb flexion/extension using a goniometer in patients with a traumatic hand injury. Two studies had a low risk of bias of which one found almost perfect reliability.

Measurements of physiological range of motion using instruments were more reliable than measurements using vision. Furthermore, measurements of physiological range of motion were also more reliable than judgements of end-feel or of accessory range of motion. We concluded that the inter-examiner reliability of passive movement assessment of joints in the upper extremity varies with the method of assessment. We recommend that clinicians measure passive physiological range of motion using goniometers or inclinometers.

Chapter 3b addresses the reliability of movement assessment of the hip, knee, and ankle-foot-toes based on studies published up to March 1, 2010. Seventeen studies were included of which seven investigated the hip, seven investigated the knee, five investigated ankle movements, and one investigated movements of the first ray of the foot. Five studies demonstrated acceptable reliability (ICC >0.75). Reliability of measurements of physiological range of motion ranged from Cohen's kappa -0.02, for rheumatologists using a goniometer to measure knee extension in patients with knee osteoarthritis, to ICC 0.97, for physiotherapists using vision to measure knee flexion in symptomatic participants. Two studies were scored as having a low risk of bias while reporting acceptable reliability for measuring physiological range of knee flexion and extension.

Measuring physiological range of knee flexion consistently yielded acceptable reliability using either vision or instruments. Judgements of end-feel were unreliable for all hip and knee movements. We concluded that the inter-examiner reliability of passive movement assessment of joints in the lower extremity joint is generally low. We recommend clinicians to be cautious when relying on these measurements for making decisions about patients with lower extremity musculoskeletal disorders.

Chapter 4 presents two studies investigating the role and position of segmental PIVM assessment within clinical reasoning and decision-making in clinical practice in manual therapy. **Chapter 4a** describes a cross-sectional, quantitative survey study. This study aimed to describe and explore the use of PIVM assessment by Dutch manual therapists and to identify factors associated with manual therapists' perception of related importance and confidence. A 13-item, structured questionnaire was developed and was sent, between September, 2006 and February, 2007, by e-mail or post to practices, individual manual therapists, and consultation platforms in the Netherlands. The e-mail response rate was 16% while the postal response rate ranged from 52% to 56%.

Three hundred and sixty-seven questionnaires were analysed; 31 (0.7%) data points were missing. Dutch manual therapists most frequently apply PIVM assessment to the cervical region and they prefer three-dimensionally coupled motions. They consider judgements of end-feel or, to a lesser extent, provocation of patient's pain as decisive for diagnostic conclusions. Respondents believe that these spinal motion tests are important for treatment decisions and they are confident in their conclusions drawn from it. We concluded that Dutch manual therapists show substantial consistency in reporting their use, interpretation, and related perceptions regarding PIVM assessment. However, their reported use and interpretation of PIVM assessment and related perceptions could only partly be substantiated by the findings from our earlier systematic review.

In the project reported in **Chapter 4b**, qualitative research methods were used to allow a deeper exploration and to improve our understanding of why, how, and when manual therapists use PIVM assessment within their clinical reasoning and decision-making in patients with spine-related disorders. Between November, 2007 and April, 2008, individual, in-depth interviews were held with eight manual therapists, leading authorities in their field and covering the range of educational programs in manual therapy in the Netherlands. Subsequently, three group interviews were conducted with consultation platforms consisting of eight to 11 manual therapists (June, 2008 to September, 2008).

From the analysis of the transcribed data, four themes emerged: contextuality, consistency, impairment orientedness, and subjectivity. These themes were interrelated and linked to various types of clinical reasoning strategies (as described by Jones et al, 2008) with professionalism (as described by Freidson, 2001) acting as a covering main theme. We found that PIVM assessment is positioned, albeit more or less routinely, as an 'add-on' test after history taking, visual inspection, and active and regional passive motion examination. In addition, our findings support a multivariable, biopsychosocial, hypothesis-oriented approach to research into manual diagnostics in general and PIVM assessment in particular.

In **Chapter 5**, a study is described that evaluated the quality of biopsychosocial history taking by manual therapists in patients with (chronic) neck or low-back pain. In Phase 1, process indicators were developed by extracting recommendations from the literature and classifying them into the three dimensions of the SCEBS method (Dutch: SCEGS methode) covering the Somatic, psychological (Cognition, Emotion, and Behaviour), and Social dimensions of chronic pain. In Phase 2, these indicators were tested in manual therapy clinical practice.

Sixty-eight literature-based recommendations were transformed into 51 process indicators. Twenty manual therapists from 27 practices contributed 108 audio recordings of history takings in patients. We concluded that manual therapists perform diagnostics of musculoskeletal pain mainly through the use of the somatic dimension of (chronic) pain. Psychological and social dimensions are inadequately covered and there is a substantial discrepancy between the actual and self-estimated use of biopsychosocial history taking. We recommend the implementation of the SCEBS method in educational programs in manual therapy.

Chapter 6 presents the protocol of a study to estimate the inter-examiner reliability among Dutch manual therapists of indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain based on a multivariable, hypothesis-oriented diagnostic reasoning and decision-making process. This study will be conducted as a repeated-measures design in which 14 pairs of manual therapists independently conduct a full diagnostic process in a consecutive series of a planned total of 165 patients presenting in primary care physiotherapy practice. Primary outcome measure is the manual therapists' decision about whether or not spinal joint mobilisations or manipulations, or both, are indicated in a patient, in isolation or as part of a multimodal treatment. The study is proposed as an initial step towards the development of an alternative approach to current classification systems and prediction rules for identifying those patients with spinal disorders that show a better response to manual therapy and which can be incorporated in randomised controlled trials.

Hypothesising that potential changes in biomechanical properties of joints after passive movements are due to immediate changes in intra-articular joint pressure and, consequently, in volume or distribution of synovial fluid, **Chapter 7** reports on a short series of basic science experiments in which ultrasonography (US) and magnetic resonance imaging (MRI) were used for visualising and measuring *in vivo* time-dependent changes in synovial fluid volume in joints of the upper cervical spine, the knee joints, and the metacarpophalangeal joints in three human subjects up to 60 minutes before and after after passive joint motion assessment, mobilisation, and high-velocity, low-amplitude thrust manipulation. We conclude that current US and MRI techniques are not appropriate for visualisation and measurement of these changes, if any, after passive movement-based joint interventions. New, innovative research is needed to generate evidence on the biomechanical effects of passive movement-based joint interventions commonly used in manual therapy.

General discussion and directives

Manual therapists strongly rely on passive intervertebral motion (PIVM) assessment for making decisions about spinal joint mobilisation and thrust manipulation in patients with neck or low-back pain.[1-4] Intervertebral motion is a physical phenomenon that can be observed and measured objectively in time and space, but it must be appreciated that the judgement of passive segmental intervertebral motion is a multidimensional construct and manual therapists conceptualise mobility and the nature of perceived resistance ('stiffness') during or at the end of a movement in an individual and subjective manner.[4;5]

Psychophysical research has shown that humans are able to discriminate between stiffness stimuli and physiotherapists showed good reliability for judging postero-anterior stiffness in spinal models.[6;7] However, in a clinical context, therapists seem to use many variables for decision-making and the reliability of stiffness or mobility judgements are negatively influenced by therapists' individual clinical experience and skill level.[8]

Our research shows that the inter-examiner reliability of segmental PIVM assessment of the cervical (motion segments C0-T4) and lumbar (T12-S1) spine is low with Cohen's kappa only rarely exceeding 0.40.[9] Our systematic review included studies published up to March 31, 2004. A search in MEDLINE (through PubMed) for additional studies published up to May 31, 2015, using the same strategy as described in Chapter 2, yielded 13 new studies[10-22] meeting the original inclusion criteria. Characteristics of these studies are described in the Appendix.

Inter-examiner reliability ranged from weighted kappa -0.26, for supine antero-posterior gliding of motion segment C0-C1, to weighted kappa 0.82, for prone postero-anterior gliding of vertebra T6. Overall, reliability was poor to moderate (kappa <0.00 to 0.60). It must be noted that the above-described 13 new studies have not been systematically searched, selected, assessed for their risk of bias and applicability of results, and analysed and there is probably a legitimate reason for a full update of our original systematic review.

We conclude that the inter-examiner reliability of segmental PIVM assessment of the cervical and lumbar spine is unacceptably low regardless the publication date of studies. We advise manual therapists not to rely solely on the outcomes of PIVM assessment when indicating spinal manual treatment in patients with neck or low-back pain, but to incorporate and integrate all clinical data from patient's history, observation, physical examination, performance tests, and questionnaires in their decision-making.

Studies based on suboptimal methods, reporting low estimates for the inter-examiner reliability of PIVM assessment, continue to be conducted and published. We believe that the idea of reaching satisfactory inter-examiner reliability for PIVM assessment should be abandoned, and the same applies to any other physical test procedure that contains a high level of subjective judgement from the examiner.

In medicine, the reliability of physical tests that require a high level of subjective judgement from the examiner is also only mediocre at best.[23] There is more than subjectivity; the complex and highly variable and personal state of health and illness over time[24] may also prove an insurmountable obstacle in the valid, unbiased estimation of the reliability of many physical examination procedures including PIVM assessment. Therefore, we propose that there is no need for new studies investigating the inter-examiner reliability of PIVM assessment within a univariable, single-test approach.

Quantitative measurements of passive physiological range of motion in extremity joints, particularly in the shoulder and knee joints, using instruments were found to have acceptable reliability.[25;26] New, high quality studies evaluating the reliability of passive range of motion assessment of the extremity joints are still relevant, as the measurements involved are less susceptible to examiner subjectivity. To further increase the value of studies and to reduce the waste of resources in such reliability research, we advise against the use of healthy participants because replication in patients will always be required regardless the initial results.

There is evidence that more experienced physiotherapists apply higher forces and also show more variation in the forces used when performing PIVM tests than students or less experienced colleagues.[27] Using higher forces negatively affects the discrimination of perceived resistance especially during the early to middle portions of the force-displacement curves of movements in symptomatic individuals.[27] In addition, studies of qualitative judgements of end-feel of segmental PIVM have almost consistently demonstrated poor inter-examiner reliability.[9;28] It is, therefore, recommended to assess PIVM by evaluating the first noticeable movement in the neutral zone behaviour of the spinal motion segment.[29]

A sustained high level of sensorimotor skill for performing PIVM assessment requires extensive training incorporating various methods to optimise tactile perception and discrimination.[29] To achieve expert performance in these skills, such training should go well beyond the initial manual therapy education and become part of a continued, self-directed engagement in ‘deliberate practice’ entailing focused training, feedback, and assessment.[30;31]

At our institute for master education in musculoskeletal physiotherapy (SOMT, Amersfoort, the Netherlands), students are trained in making decisions about diagnoses and treatment indications by integrating their acquired skills for performing and interpreting PIVM assessment with all other information from the clinical encounter. Reflective case discussions with peer observation and review in small groups, where possible also including direct input from patients, are used for this integrative training of analytic clinical reasoning strategies.[32;33] In accordance with national guidelines[34], students learn to search for consistency between impairments, activity limitations, and participation restrictions within clinical data while accounting for personal and external factors. Finding such consistency would imply, for example, that reducing impairments in spinal segmental mobility, as detected by PIVM assessment, through the application of spinal manual therapy such as mobilisation or manipulation will linearly lead to an improvement in activity levels and participation.

Living systems, and human disease for that matter, are, however, highly complex and behave unpredictably with many interdependent variables interacting non-linearly.[35] For instance, in non-specific neck and low-back pain, associations between spinal range of motion on the one hand and pain and disability on the other are weak or non-existent[36-41] making judgements about consistency between clinical data extremely difficult.

Students of manual therapy, as well as more experienced therapists, need to be confronted continually with the clinical uncertainties surrounding the subjective judgements about motion impairments, the mechanisms and effectiveness of joint mobilisation or manipulation, the complex relationships between clinical variables, and the value and interpretation of patient-reported outcomes. Nowadays, education at the masters level enables manual therapists to learn to deal with these uncertainties by developing the knowledge, skills, and attitudes for an advanced level of evidence-based clinical reasoning, the justification of their decisions, a critical approach to practice, and a high level of self-analysis.[42] In the Netherlands, a necessary next step is to elevate the manual therapy profession to the academic level by offering university education.

Neck and low-back pain, as well as many other musculoskeletal disorders, have since decades been among the leading causes of disability worldwide.[43] Unfortunately, patients with these conditions still seem to fare hardly better than natural course on an excess of treatment interventions available to them including spinal manual therapy.[44;45] The manual therapy profession is now at the crossroads where it can either show itself to be a legitimate partner for patients and societies building on new, innovative research designs, or fail to do so.

Our research supports a multivariable, biopsychosocial, hypothesis-oriented approach to evaluating clinical diagnostics and decision-making in manual therapy as opposed to continuing investigating the value of single tests.[3;4] We propose methods to incorporate such an approach into reliability studies and randomised controlled trials with the aim to, eventually, better identify those patients responding (or not) to spinal manual therapy.[46]

Finally, I return to the question stemming from the early beginnings of my clinical career and eventually inspiring this thesis: Could my uncertainty when performing and interpreting PIVM assessment, as compared to my confident teachers, clinical mentors, and colleagues, be attributed to my incompetence or to their overconfidence? I have no doubt that I was then, and most likely still am, not properly skilled to assess segmental PIVM at the required high expertise level. After all, I still have not made my 10.000 hours of ‘deliberate practice’.[47] On the other hand, experienced manual therapists tend to overestimate their performance.[3;48] Fortunately, a certain level of overconfidence is regarded as a general feature of human behaviour that can be favourable when definitive decisions or actions are demanded.[49] If nothing else, this at least shows that manual therapists are in fact human!

References

1. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.
2. MacDermid JC, Walton DM, Côté P, Santaguida PL, Gross A, Carlesso L, and ICON. Use of outcome measures in managing neck pain: An international multidisciplinary survey. Open Orthop J 2013;7:506-20.
3. Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine: A survey among physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51.
4. Van Trijffel E, Plochg T, Van Hartingsveld F, Lucas C, Oostendorp RAB. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual therapy. A qualitative interview study. J Man Manip Ther 2010;18:111-8.
5. Maher CG, Simmonds M, Adams R. Therapists' conceptualization and characterization of the clinical concept of spinal stiffness. Phys Ther 1998;78:289-300.
6. Maher C, Adams R. Is the clinical concept of spinal stiffness multidimensional? Phys Ther 1995;75:854-60.
7. Maher C, Adams R. A psychophysical evaluation of manual stiffness discrimination. Aust J Physiother 1995;41:161-7.
8. McGill S. Invited commentary on intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. Arch Phys Med Rehabil 2013;94:1635-7.
9. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69.
10. Bracht MA, Nunes GS, Celestino J, Schwertner DS, França LC, De Noronha M. Inter- and intra-observer agreement of the motion palpation test for lumbar vertebral rotational asymmetry. Physiother Canada 2015;67:169-73.
11. Brismée J-M, Atwood K, Fain M, Hodges J, Sperle A, Swaney M, Phelps V, Van Paridon D, Matthijs O, Sizer P. Interrater reliability of palpation of three-dimensional segmental motion of the lumbar spine. J Man Manip Ther 2005;13:215-20.

12. Brismée J-M, Gipson D, Ivie D, Lopez A, Moore M, Matthijs O, Phelps V, Sawyer S, Sizer P. Interrater reliability of a passive physiological intervertebral motion test in the mid-thoracic spine. J Manipulative Physiol Ther 2006;29:368-73.
13. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 2006;87:1388-95.
14. Deore M, May S. The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. J Bodywork Movement Ther 2012;16:289-93.
15. Hanney WJ, George SZ, Kolber MJ, Young I, Salamh PA, Cleland JA. Inter-rater reliability of select physical examination procedures in patients with neck pain. Physiother Theory Pract 2014;30:345-52.
16. Johansson F. Interexaminer reliability of lumbar segmental mobility tests. Man Ther 2006;11:331-6.
17. Lundberg G, Gerdle B. The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel. Scand J Rehabil Med 1999;31:197-206.
18. Manning DM, Dedrick GS, Sizer PS, Brismée J-M. Reliability of a seated three-dimensional passive intervertebral motion test for mobility, end-feel, and pain provocation in patients with cervicalgia. J Man Manip Ther 2012;20:135-41.
19. Minaya Muñoz FJ, Valera Garrido F, Veiga Monasterioguren X. Fiabilidad interobservador del test de movilidad pasiva accesoria intervertebral lumbar y su correlación con la galvanopalpación en sujetos sanos. Rev Iberoam Fisioter Kinesiol 2010;13:4-9.
20. Ogince M, Hall T, Robinson K, Blackmore AM. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headache. Man Ther 2007;12:256-62.
21. Piva SR, Erhard RE, Childs JD, Browder DA. Inter-tester reliability of passive intervertebral and active movements of the cervical spine. Man Ther 2006;11:321-30.
22. Schreiner M. Interrater-Reliabilität passiver physiologischer intervertebraler Bewegungen in der Sagittalebene der LWS. Manuelletherapie 2008;12:201-5.
23. Joshua AM, Celermajer DS, Stockler MR. Beauty is in the eye of the examiner: Reaching agreement about physical signs and their value. Intern Med J 2005;35:178-87.

24. Sturmberg JP, Martin CM, Katerndahl DA. Systems and complexity thinking in general practice literature: An integrative, historical narrative review. Ann Fam Med 2014;12:66-74.
25. Van de Pol RJ, Van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: A systematic review. J Physiother 2010;56:7-17.
26. Van Trijffel E, Van de Pol RJ, Oostendorp RAB, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review. J Physiother 2010;56:223-35.
27. Hazle CR Jr, Nitz AJ. A simulated passive intervertebral motion task: Observations of performance in a cross-sectional study. J Man Manip Ther 2012;20:121-9.
28. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
29. Nyberg RE, Smith AR Jr. The science of spinal motion palpation: A review and update with implications for assessment and intervention. J Man Manip Ther 2013;21:160-7.
30. Davis DA. How to help professionals maintain and improve their knowledge and skills: Triangulating best practices in medicine. In: Ericsson KA (editor). Development of professional expertise. Toward measurement of expert performance and design of optimal learning environments. New York, NY: Cambridge University Press, 2009.
31. Ericsson KA. Enhancing the development of professional performance: Implications from the study of deliberate practice. In: Ericsson KA (editor). Development of professional expertise. Toward measurement of expert performance and design of optimal learning environments. New York, NY: Cambridge University Press, 2009.
32. Greenhalgh T, Howick J, Maskrey N, for the Evidence Based Medicine Renaissance Group. Evidence based medicine: A movement in crisis? BMJ 2014;348:g3725.
33. O'Dunn-Orto A, Hartling L, Campbell S, Oswald AE. Teaching musculoskeletal skills to medical trainees and physicians: A Best Evidence in Medical Education systematic review of strategies and their effectiveness: BEME Guide No. 18. Med Teach 2012;34:93-102.
34. Heemskerk MABM, Staal JB, Bierma-Zeinstra SMA, De Haan G, Hagenaars LHA, Lanser K, Van der Windt DAWM, Oostendorp RAB, Hendriks HJM. KNGF Guideline Complaints of Arms, Neck and or Shoulder, 2010. Available at: https://www.fysionet-

evidencebased.nl/images/pdfs/richtlijnen/kans_2010/kans_praktijkrichtlijn.pdf. [In Dutch] (accessed 24 November, 2014)

35. Mossman K. The complexity paradox. The more answers we find, the more questions we have. New York, NY: Oxford University Press, 2014.
36. Leboeuf-Yde C, Van Dijk J, Franz C, Hustad SA, Olsen D, Pihl T, Röbech R, Vendrup SS, Bendix T, Kyvik KO. Motion palpation findings and self-reported low back pain in a population-based study sample. J Manipul Physiol Ther 2002;25:80-7.
37. Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. Clin Rehabil 1999;13:211-8.
38. Olson SL, O'Connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. J Orthop Sports Phys Ther 2000;30:13-20.
39. Parks KA, Crichton KS, Goldford RJ, McGill SM. A comparison of lumbar range of motion and functional ability scores in patients with low back pain: Assessment for range of motion validity. Spine 2003;28:380-4.
40. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 1998;78:951-63.
41. Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. Phys Ther 2000;80:240-50.
42. Rushton A, Lindsay G. Defining the construct of masters level clinical practice in manipulative physiotherapy. Man Ther 2008;15:93-9.
43. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163-96.
44. Haldeman S, Dagenais S. A supermarket approach to the evidence-informed management of chronic low back pain. Spine J 2008;8-1-7.
45. Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: The HUNT study. Pain 2013;154:1237-44.
46. Van Trijffel E, Lindeboom R, Bossuyt PMM, Schmitt MA, Lucas C, Koes BW, Oostendorp R. Indicating spinal joint mobilisations or manipulations in patients with neck or low-back

pain: Protocol of an inter-examiner reliability study among manual therapists. Chiropr Man Therap 2014;22:22.

47. Ericsson KA. Expertise. Curr Biol 2014;24:R508-10.
48. Oostendorp RAB, Elvers H, Mikojalewska E, Laekeman M, Van Trijffel E, Samwel H, Duquet W. Manual physical therapists' use of biopsychosocial history taking in the management of patients with back or neck pain in clinical practice. ScientificWorldJournal 2015;2015:170463.
49. Croskerry P, Norman G. Overconfidence in clinical decision making. Am J Med 2008;121:S24-9.

Appendix

Studies published between March 31, 2004 and May 31, 2015*

Study (First author)	**Judgement criteria and scales**	**Position and movement**	**Reliability**
Bracht[10]	Asymmetry in rotational movement (left-right comparison)	*Subject prone:* unilateral P-A glide left and right (via transverse processes)	L2: κ 0.12 to 0.23
Brismeé[11]	Direction of lateral flexion leading to greatest rotation (ipsilateral or contralateral)	*Subject left side-lying:* right rotation in left and right lateral flexion with hips and knees in 45° and 70°, respectively	L4-L5: κ -0.08 to 0.04
Brismeé[12]	Movement pattern leading to greatest rotation (ipsilateral or contralateral)	*Subject seated:* rotation in extension and left and right lateral flexion	T6-T7: κ 0.27 to 0.65
Cleland[13]	Normal mobility-hypermobile-hypomobile (left-right comparison or relative to segments above and below the tested level)	*Subject supine:* A-P glide C0-C1 rotation C1-C2 *Subject prone:* P-A glide C2-T9	C0-C1: κ_w -0.26, 0.46 C1-C2: κ_w 0.72, 0.74 C2 to C7: κ_w 0.01 to 0.54 T1 to T9: κ_w 0.13 to 0.82
Deore[14]	Hypomobile-normal-hypermobile	*Subject prone:* P-A glide	L1: κ 0.20 L2: κ 0.19 L3: κ 0.03 L4: κ 0.00 L5: κ 0.01
Hanney[15]	Hypomobile-normal-hypermobile	*Subject supine:* A-P glide C0-C1 rotation C1-C2 lateral glide C2-T9	C0-C1: κ_w 0.15 C1-C2: κ_w 0.31 C2-C3: κ_w 0.30 C3-C4: κ_w 0.22 C4-C5: κ_w 0.43 C5-C6: κ_w 0.30 C6-C7: κ_w 0.23
Johansson[16]	Extreme hypomobility-hypomobility-normal mobility-hypermobility-extreme hypermobility	*Subject side-lying:* flexion, extension with hips and knees slightly flexed	L1-L2: κ_w -0.06 to 0.27 L2-L3: κ_w -0.14 to 0.54 L3-L4: κ_w -0.12 to 0.28 L4-L5: κ_w 0.02 to 0.17 L5-S1: κ_w -0.09 to 0.56
Lundberg[17]	Extreme hypomobility-moderate hypomobility-normal mobility-moderate hypermobility-extreme	*Subject side-lying:* overall segmental mobility of flexion, extension, left and right rotation, translatoric joint	T10-T11: κ_w NA T11-T12: κ_w 0.73 T12-L1: κ_w 0.59

	hypermobility	play with hips and knees flexed	L1-L2: κ_w 0.68 L2-L3: κ_w 0.61 L3-L4: κ_w NA L4-L5: κ_w 0.75 L5-S1: κ_w 0.70
Manning[18]	Hypomobility and hard end-feel as compared to the segment above the tested level	*Subject seated:* three-dimensional extension and ipsilateral lateral flexion and rotation	C2-C3: *hypomobility* κ 0.32, 0.33 and *end-feel* κ 0.28, 0.37 C3-C4: κ 0.33, 0.41 and κ 0.32, 0.50 C4-C5: κ 0.41, 0.48 and κ 0.39, 0.47 C5-C6: κ 0.21, 0.57 and κ 0.25, 0.60 C6-C7: κ 0.22, 0.58 and κ 0.28, 0.59
Minaya Muñoz[19]	Abnormal end-feel	*Subject prone:* unilateral P-A glide left and right	L1-L2: PA *right* 86.6%, *left* 86.6% L2-L3: PA 83.3%, 83.3% L3-L4: PA 90.0%, 90.0% L4-L5: PA 86.6%, 93.3% L5-S1: PA 86.6%, 76.6% Overall κ 0.50
Ogince[20]	Limited range of motion based on a firm end-feel and therapist's interpretation of a 10° reduction	*Subject supine:* left and right rotation with cervical spine fully flexed	C1-C2: κ 0.81
Piva[21]	Normal mobility-hypomobile	*Subject supine:* lateral glide C0-C1 lateral displacement C1 left and right rotation with cervical spine fully flexed lateral glide C2-C6	C0-C1: κ 0.81 C1: κ 0.35 C1-C2: κ 0.30 C2: κ 0.46 C3: κ 0.25 C4: κ 0.27 C5: κ 0.18 C6: κ -0.07
Schreiner[22]	Normal mobility-hypermobile-hypomobile	*Subject side-lying:* flexion, extension with hips and knees flexed	T12-L1: κ *extension* -0.11, *flexion* κ NA L1-L2: κ NA, κ NA L2-L3: κ 0.00, κ -0.07 L3-L4: κ 0.29, κ -0.04 L4-L5: κ -0.24, κ -0.14 L5-S1: κ 0.02, κ -0.02

A-P: antero-posterior, κ: kappa, κ_w: weighted kappa, NA: not available, PA: percentage agreement, P-A: postero-anterior

*Except Lundberg: published before March 31, 2004

Samenvatting

Role and reliability of passive joint motion assessment: Towards multivariable diagnostics and decision-making in manual therapy

Samenvatting

Vanuit wetenschappelijk perspectief bestaat onzekerheid over het gebruik en de waarde van het passief uitgevoerde intervertebraal bewegingsonderzoek van de wervelkolom binnen de diagnostiek en klinische besluitvorming in de manuele therapie. Tegen die achtergrond beschrijft **Hoofdstuk 1** de twee hoofddoelstellingen van het onderzoek waarover wordt gerapporteerd in dit proefschrift: (1) het evalueren van de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde gewrichtsonderzoek van de wervelkolom en de extremiteiten en (2) het inventariseren van de rol en positie van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek van de wervelkolom in het proces van klinisch redeneren en de klinische besluitvorming in de klinische praktijk van de manuele therapie bij patiënten met wervelkolomgerelateerde aandoeningen.

Hoofdstuk 2 beschrijft een systematische literatuurstudie naar de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek van de cervicale (bewegingssegmenten C0-T4) en lumbale (T12-S1) wervelkolom. Negentien onderzoeken gepubliceerd vóór 31 maart 2004 werden geïncludeerd waarvan negen de resultaten van het bewegingsonderzoek van de cervicale wervelkolom en 10 die van de lumbale wervelkolom beschreven. De waarden van de tussenbeoordelaarsbetrouwbaarheid varieerden van Cohen's kappa -0,32, voor chiropractisch onderzoek van de beweeglijkheid van lateroflexie en rotatie van het bewegingssegment T3-T4 (in zit) bij patiënten met niet-musculoskeletale aandoeningen, tot een gewogen kappa 0,75, voor fysiotherapeuten die het bewegingssegment L5-S1 onderzochten in angulaire en translatorische bewegingsrichtingen (in zijligging) bij patiënten met lage-rugpijn. De waarden van kappa varieerden in de verschillende onderzoeken van <0,00 tot 0,40, oftewel van 'slecht' tot 'redelijk'. Drie onderzoeken hadden een laag risico op bias waarvan er één redelijke tot behoorlijke betrouwbaarheid (kappa 0,28 tot 0,43) vond voor beoordelingen van stijfheid van de bewegingssegmenten C1-C2, C2-C3, C7-T1 en van de eerste rib. Het bewegingsonderzoek van de bewegingssegmenten C1-C2 en C2-C3 liet consistent minimaal redelijke betrouwbaarheid (kappa >0,21) zien. De beoordelingen van de kwantiteit en de kwaliteit van beweging waren even (on)betrouwbaar. Wij concludeerden dat de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek van de cervicale en lumbale wervelkolom laag is evenals de methodologische kwaliteit van de geïncludeerde onderzoeken.

Hoofdstuk 3 presenteert twee systematische literatuurstudies naar de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde bewegingsonderzoek van de gewrichten van de bovenste en onderste extremiteiten. Wij gebruikten vergelijkbare methoden als in Hoofdstuk 2. **Hoofdstuk 3a** betreft de betrouwbaarheid van het bewegingsonderzoek van de schouder, de elleboog en de pols, hand en vingers uit onderzoeken die zijn gepubliceerd vóór 1 juli 2009. Eenentwintig onderzoeken werden geïncludeerd waarvan er 13 het bewegingsonderzoek van de schouder evalueerden, twee dat van de elleboog, vier dat van de polsbewegingen, één dat van de bewegingen van de vingergewrichten en één dat van de duimbewegingen. Elf onderzoeken lieten acceptabele betrouwbaarheid zien (ICC >0,75). De betrouwbaarheid varieerde sterk met de methode van het meten van de bewegingsuitslagen en de waarden van de ICC varieerden van 0,26 (95% BI -0,01 tot 0,69), voor het visueel meten van de fysiologische bewegingsuitslag van de endorotatie van de schouder bij patiënten, tot 0,99 (95% BI 0,98 tot 1,0), voor het meten van de passieve *torque-controlled* fysiologische bewegingsuitslag van de flexie en extensie van de duim en de vingers met een goniometer bij patiënten met een posttraumatisch handletsel. Twee onderzoeken hadden een laag risico op bias waarvan er één bijna perfecte betrouwbaarheid liet zien.

Metingen van passief uitgevoerde fysiologische bewegingsuitslagen van gewrichten met behulp van instrumenten waren meer betrouwbaar dan die met visuele metingen. Bovendien waren de metingen van passief uitgevoerde fysiologische bewegingsuitslagen meer betrouwbaar dan de beoordelingen van het eindgevoel of van translatorische bewegingsuitslagen. Wij concludeerden dat de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde bewegingsonderzoek van gewrichten van de bovenste extremiteit varieert met de methode van meten. Wij bevelen clinici aan om bij het meten van passief uitgevoerde fysiologische bewegingsuitslagen van deze gewrichten gebruik te maken van goniometers of inclinometers.

Hoofdstuk 3b betreft de betrouwbaarheid van het passief uitgevoerde bewegingsonderzoek van de heup, de knie en de enkel, voet en tenen uit onderzoeken die zijn gepubliceerd vóór 1 maart 2010. Zeventien onderzoeken werden geïncludeerd waarvan zeven het bewegingsonderzoek van de heup onderzochten, zeven dat van de knie, vijf dat van de enkelbewegingen en één dat van de bewegingen van de eerste straal van de voet. Vijf onderzoeken lieten acceptabele betrouwbaarheid zien (ICC >0,75). De waarden van de betrouwbaarheid van de metingen van de passief uitgevoerde fysiologische bewegingsuitslagen varieerden van Cohen's kappa -0,02, voor het meten van de extensie van de knie met behulp van een goniometer door reumatologen bij patiënten met knieartrose, tot ICC 0,97, voor het visueel meten van de flexie van de knie door fysiotherapeuten bij symptomatische deelnemers. Twee onderzoeken hadden een laag risico op bias en rapporteerden acceptabele betrouwbaarheid voor het meten van de fysiologische bewegingsuitslag van de flexie en extensie van de knie.

Het meten van de passief uitgevoerde fysiologische bewegingsuitslag van de flexie van de knie leidde consistent tot acceptabele betrouwbaarheid ongeacht de methode van meten. Beoordelingen van het eindgevoel van alle bewegingen van de heup en de knie waren onbetrouwbaar. Wij concludeerden dat de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde bewegingsonderzoek van de gewrichten van de onderste extremiteit in het algemeen laag is. Wij bevelen clinici aan om terughoudend te zijn met het op dit bewegingsonderzoek baseren van beslissingen voor patiënten met musculoskeletale aandoeningen van de onderste extremiteit.

Hoofdstuk 4 presenteert twee onderzoeken naar de rol en positie van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek binnen het klinisch redeneren en de klinische besluitvorming in de klinische praktijk van de manuele therapie. **Hoofdstuk 4a** beschrijft een cross-sectioneel, kwantitatief surveyonderzoek. Het doel van dit onderzoek was om het gebruik van het intervertebraal bewegingsonderzoek door Nederlandse manueeltherapeuten te beschrijven en te exploreren en om factoren te identificeren die zijn geassocieerd met de perceptie van manueeltherapeuten over het aan dit bewegingsonderzoek gerelateerde belang en vertrouwen. Een gestructureerde lijst met 13 vragen werd ontwikkeld en tussen september 2006 en februari 2007 verstuurd per e-mail of per post naar praktijken, individuele manueeltherapeuten en intercollegiale overleggroepen

(IOF’s) in Nederland. De respons via e-mail bedroeg 16% en die per post varieerde van 52% tot 56%.

Driehonderd en zevenenzestig vragenlijsten werden geanalyseerd; er waren 31 (0,7%) missende waarden. Nederlandse manueeltherapeuten gebruiken het passief uitgevoerde intervertebraal bewegingsonderzoek het vaakst aan de cervicale wervelkolom en zij hebben een voorkeur voor driedimensionale gekoppelde bewegingen. Zij beschouwen het eindgevoel of, in mindere mate, provocatie van de pijn van de patiënt als doorslaggevend voor diagnostische conclusies. De respondenten waren van mening dat deze bewegingstests voor de wervelkolom belangrijk zijn voor hun behandelbeslissingen en zij hebben vertrouwen in de conclusies die zij uit deze tests trekken. Wij concludeerden dat Nederlandse manueeltherapeuten onderling consistent waren in het rapporteren van hun gebruik, interpretatie en gerelateerde percepties rondom het passief uitgevoerde intervertebraal bewegingsonderzoek. Hun gebruik, interpretatie en percepties van dit bewegingsonderzoek konden echter slechts ten dele worden ondersteund door de bevindingen uit de eerder door ons uitgevoerde systematische literatuurstudie.

In het project waarover in **Hoofdstuk 4b** wordt gerapporteerd, werden kwalitatieve onderzoeksmethoden gebruikt voor een diepere exploratie en een beter begrip van waarom, hoe en wanneer manueeltherapeuten het passief intervertebraal bewegingsonderzoek gebruiken binnen hun klinisch redeneren en klinische besluitvorming bij patiënten met wervelkolomgerelateerde aandoeningen. Tussen november 2007 en april 2008 werden individuele diepte-interviews gehouden met acht manueeltherapeuten die werden beschouwd als autoriteiten namens de verschillende opleidingen manuele therapie in Nederland. Vervolgens werden drie groepsinterviews gehouden met IOF’s bestaande uit acht tot 11 manueeltherapeuten (juni 2008 - september 2008).

Uit de analyse van de getranscribeerde data kwamen vier thema’s naar voren: contextualiteit, consistentie, stoornisgerichtheid en subjectiviteit. Deze thema’s waren onderling gerelateerd en werden gekoppeld aan diverse vormen van klinisch redeneren (zoals beschreven door Jones et al, 2008) waarbij professionalisme (zoals beschreven door Freidson, 2001) fungeerde als overkoepelend hoofdthema. Wij concludeerden dat het passief uitgevoerde intervertebraal bewegingsonderzoek is gepositioneerd, zij het min of meer routinematig, als een ‘*add-on*’-test na de anamnese, de visuele inspectie en het actief en regionaal passief uitgevoerde bewegingsfunctieonderzoek. Onze bevindingen ondersteunen een multivariabele, biopsychosociale, hypothese-gestuurde benadering van onderzoek naar manuele diagnostiek en naar het passief uitgevoerde intervertebraal bewegingsonderzoek in het bijzonder.

In **Hoofdstuk 5** wordt een onderzoek beschreven waarin de kwaliteit werd geëvalueerd van de biopsychosociale anamnese door manueeltherapeuten bij patiënten met (chronische) nekpijn of lage-rugpijn. In Fase 1 werden procesindicatoren ontwikkeld vanuit aanbevelingen in de literatuur. Deze indicatoren werden vervolgens geklasseerd volgens de dimensies van de SCEGS methode die de Somatische, de psychologische (Cognitie, Emotie en Gedrag) en de Sociale dimensies van chronische pijn beslaat. In Fase 2 werden de ontwikkelde procesindicatoren getest in de klinische praktijk van de manuele therapie.

Achtenzestig aanbevelingen uit de literatuur werden omgezet in 51 procesindicatoren. Tweeëntwintig manueeltherapeuten uit 27 praktijken leverden 108 geluidsopnames van anamneses bij patiënten. Wij concludeerden dat manueeltherapeuten het bevragen van patiënten met musculoskeletale pijn vooral uitvoeren binnen de somatische dimensie van (chronische) pijn. De psychologische en sociale dimensies worden onvoldoende gedekt en er bestaat een sterke discrepantie tussen de daadwerkelijke en de zelf-geschatte uitvoering van de biopsychosociale anamnese. Wij bevelen aan om de SCEGS methode te implementeren in de onderwijsprogramma’s in de manuele therapie.

Hoofdstuk 6 beschrijft het protocol van een onderzoek naar het schatten van de tussenbeoordelaarsbetrouwbaarheid onder Nederlandse manueeltherapeuten van het stellen van een indicatie voor mobilisatie of manipulatie van de wervelkolom bij patiënten met nekpijn of lage-rugpijn op grond van een multivariabel, hypothese-gestuurd diagnostisch proces. Dit onderzoek kent een *repeated-measures* opzet waarbij 14 paren van manueeltherapeuten onafhankelijk van elkaar bij een gepland aantal van 165 opeenvolgende patiënten die zich presenteren in de eerstelijns fysiotherapiepraktijk een volledig diagnostisch proces uitvoeren. De primaire uitkomstmaat is het besluit van de manueeltherapeut bij elke patiënt of mobilisatie of manipulatie van de wervelkolom, of beide, zijn geïndiceerd, op zichzelf of als onderdeel van een multimodale behandelinterventie. Het onderzoek vormt een initiële stap naar het ontwikkelen van een alternatieve wijze, tegenover de huidige classificaties en predictieregels, voor het identificeren van patiënten die reageren op manuele therapie aan de wervelkolom en die kan worden geïncorporeerd in gerandomiseerde klinische effectstudies.

Uitgaande van de hypothese dat mogelijke veranderingen in de biomechanische eigenschappen van gewrichten na passief uitgevoerde bewegingen het gevolg zijn van onmiddelijke veranderingen in de intra-articulaire druk en vervolgens in het volume of de verdeling van de synoviale vloeistof, rapporteert **Hoofdstuk 7** over een korte serie van fundamentele onderzoeksexperimenten waarin ultrasonografie (US) en magnetic resonance imaging (MRI) zijn gebruikt voor het visualiseren en meten van *in vivo* veranderingen in de tijd in het volume van de synovia in gewrichten van de hoog-cervicale wervelkolom, de kniegewrichten en de metacarpophalangeale gewrichten bij drie proefpersonen tot 60 minuten voor en na het passief uitgevoerde bewegingsonderzoek, gewrichtsmobilisaties en –manipulaties. Wij concludeerden dat de huidige technieken voor US en MRI niet zijn geschikt voor het visualiseren en meten van deze veranderingen, indien al aanwezig, na passief uitgevoerde bewegingsinterventies aan gewrichten. Nieuw, innovatief onderzoek is noodzakelijk om evidence te genereren over de biomechanische effecten van passief uitgevoerde gewrichtsinterventies zoals die worden toegepast in de manuele therapie.

Algemene discussie en richtingbepaling

Manueeltherapeuten laten zich sterk leiden door het passief uitgevoerde intervertebraal bewegingsonderzoek bij hun besluitvorming over het toepassen van gewrichtsmobilisaties en –manipulaties van de wervelkolom bij patiënten met nekpijn of lage-rugpijn.[1-4] Intervertebraal bewegen is een fysisch verschijnsel dat in tijd en plaats kan worden waargenomen en gemeten, maar het moet worden onderkend dat de beoordeling van de passieve segmentale intervertebrale beweeglijkheid een multidimensioneel construct is. Manueeltherapeuten conceptualiseren die beweeglijkheid en de aard van de ervaren weerstand (‘stijfheid’) tijdens de beweging of op de bewegingsgrens namelijk op een individuele en subjectieve wijze.[4;5]

Onderzoek uit de psychofysica heeft laten zien dat mensen in staat zijn om te discrimineren tussen stijfheidsstimuli en fysiotherapeuten toonden goede betrouwbaarheid bij het beoordelen van de voor-achterwaartse stijfheid in wervelkolommodellen.[6;7] In een klinische context echter lijken therapeuten vele variabelen te gebruiken in hun besluitvorming en de betrouwbaarheid van beoordelingen over stijfheid of beweeglijkheid wordt ongunstig beïnvloed door de individuele klinische ervaring en het individuele vaardigheidsniveau van therapeuten.[8]

Ons onderzoek liet zien dat de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek van de cervicale (bewegingssegmenten C0-T4) en lumbale (T12-S1) wervelkolom laag is met waarden voor Cohen's kappa die zelden de 0,40 overstegen.[9] Onze systematische literatuurstudie bevatte onderzoeken gepubliceerd tot 31 maart 2004. Een zoektocht in MEDLINE (via PubMed) voor aanvullende onderzoeken gepubliceerd tot 31 mei 2015, gebruikmakend van dezelfde zoekstrategie zoals beschreven in Hoofdstuk 2, leverde 13 nieuwe onderzoeken[10-22] op die voldeden aan de oorspronkelijke inclusiecriteria. De kenmerken van deze onderzoeken zijn beschreven in de Appendix.

De waarden van de tussenbeoordelaarsbetrouwbaarheid varieerden van gewogen kappa -0,26, voor het antero-posterior glijden van bewegingssegment C0-C1 (in rugligging), tot gewogen kappa 0,82, voor het postero-anterior glijden van wervel T6 (in buikligging). Over het geheel genomen was de betrouwbaarheid slecht tot behoorlijk (kappa <0,00 to 0,60). Hierbij wordt aangetekend dat de genoemde 13 nieuwe onderzoeken niet op een systematische wijze zijn gezocht, geselecteerd, beoordeeld op hun risico op bias en toepasbaarheid van de resultaten en geanalyseerd, hetgeen een geldige aanleiding kan zijn voor een volledige update van onze oorspronkelijke systematische literatuurstudie.

Wij concluderen dat de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde segmentaal intervertebraal bewegingsonderzoek van de cervicale en lumbale wervelkolom onaanvaardbaar laag is ongeacht de publicatiedatum van de onderzoeken. Wij adviseren manueeltherapeuten om indicaties voor de manueeltherapeutische behandeling van patiënten met nekpijn of lage-rugpijn niet alleen te baseren op de uitkomsten van het passief uitgevoerde intervertebraal bewegingsonderzoek, maar om alle klinische informatie uit anamnese, observatie, lichamelijk onderzoek, performance tests en vragenlijsten te incorporeren en te integreren in hun besluitvorming.

Onderzoeken die gebruikmaken van suboptimale methoden en die daarbij lage schatters voor de tussenbeoordelaarsbetrouwbaarheid van het passief uitgevoerde intervertebraal bewegingsonderzoek laten zien, worden nog steeds uitgevoerd en gepubliceerd. Wij zijn van mening dat het streven naar een bevredigende tussenbeoordelaarsbetrouwbaarheid van het passief intervertebraal bewegingsonderzoek moet worden verlaten en hetzelfde geldt voor elke andere testprocedure uit het lichamelijk onderzoek die een hoge mate van subjectiviteit bevat in de klinische beoordeling door de (manueel)therapeut.

Ook in de geneeskunde is de betrouwbaarheid van lichamelijke tests met een hoge mate van subjectiviteit in het oordeel door de arts ten hoogste middelmatig.[23] Er is meer dan subjectiviteit; de complexe en individueel sterk variabele toestand van gezondheid en ziekte in de tijd[24] kan eveneens een onoverbrugbaar obstakel blijken te vormen bij het valide, onvertekend schatten van de betrouwbaarheid van vele lichamelijke testprocedures waaronder het passief intervertebraal bewegingsonderzoek. Wij stellen daarom dat er geen behoefte bestaat aan nieuwe studies die de tussenbeoordelaarsbetrouwbaarheid van het passief intervertebraal bewegingsonderzoek evalueren binnen een univariabele, *single-test* benadering.

Kwantitatieve metingen van passieve fysiologische bewegingsuitslagen in gewrichten van de extremiteiten, in het bijzonder in de gewrichten van de schouder en de knie, met behulp van instrumenten bleken acceptabele betrouwbaarheid te vertonen.[25;26] Nieuwe onderzoeken van goede methodologische kwaliteit die de betrouwbaarheid van het passief onderzoek van bewegingsuitslagen van gewrichten van de extremiteiten evalueren zijn relevant, mede aangezien de metingen van deze bewegingsuitslagen minder ontvankelijk zijn voor subjectiviteit aan de kant van de beoordelaar. Om in zulk betrouwbaarheidsonderzoek de waarde van het onderzoek te vergroten en de verspilling van middelen te reduceren, adviseren wij om het onderzoek niet uit te voeren met gezonde deelnemers. Replicatie met patiënten zal namelijk altijd noodzakelijk zijn ongeacht de initiële resultaten.

Er zijn aanwijzingen dat ervaren manueeltherapeuten grotere krachten gebruiken en ook meer variatie vertonen in krachten tijdens het uitvoeren van het passief intervertebraal bewegingsonderzoek in vergelijking met studenten of minder ervaren collega's.[27] Het gebruik van grotere krachten heeft een negatief effect op de discriminatie van de ervaren weerstand vooral tijdens de vroege en middelste gedeeltes van de kracht-verlengingscurves van bewegingen bij symptomatische personen.[27] Daarnaast hebben onderzoeken naar kwalitatieve beoordelingen van het eindgevoel van passief uitgevoerde segmentale intervertebrale bewegingen bijna consistent slechte tussenbeoordelaarsbetrouwbaarheid laten zien.[9;28] Het wordt om die redenen aanbevolen om passieve intervertebrale bewegingen te beoordelen op de eerst waarneembare beweging in de neutrale zone van het spinale bewegingssegment.[29]

Een duurzaam hoog niveau van sensomotorische vaardigheid voor het uitvoeren van het passief intervertebraal bewegingsonderzoek vereist extensieve training waarin diverse methoden voor het optimaliseren van tactiele perceptie en discriminatie zijn opgenomen.[29] Voor het bereiken van het niveau van expert in deze vaardigheden moet die training nadrukkelijk verder gaan dan de initiële scholing in manuele therapie en onderdeel worden van een levenslang, zelfgestuurd traject van '*deliberate practice*' dat doelgerichte training, feedback en toetsing omvat.[30;31]

Binnen ons instituut voor masteronderwijs in musculoskeletale fysiotherapie (SOMT, Amersfoort) worden studenten getraind in de besluitvorming tijdens het diagnostisch proces en het stellen van behandelindicaties door hun verworven vaardigheden voor het uitvoeren en interpreteren van het passief intervertebraal bewegingsonderzoek te integreren met alle overige klinische informatie. Voor deze integratieve training van analytische strategieën voor het klinisch redeneren worden reflectieve besprekingen van casuïstieken gebruikt, waarbij studenten in kleine groepen elkaar observeren en beoordelen met, waar mogelijk, de directe input van patiënten.[32;33] In overeenstemming met nationale praktijkrichtlijnen[34] leren studenten te zoeken naar consistentie tussen stoornissen, beperkingen in activiteiten en participatieproblemen in de klinische informatie daarbij rekening houdend met persoonlijke en externe factoren. Het vinden van een dergelijke consistentie zou kunnen impliceren dat het verminderen van stoornissen in de beweeglijkheid van spinale bewegingssegmenten, geïdentificeerd met het passief uitgevoerde intervertebraal bewegingsonderzoek, door het toepassen van mobilisatie of manipulatie aan de wervelkolom leidt tot een evenredige verbetering van het activiteitenniveau en de participatie.

Levende systemen, en menselijke ziekten in het bijzonder, zijn echter zeer complex en onvoorspelbaar met vele onafhankelijke variabelen die op non-lineaire wijze met elkaar interacteren.[35] In het geval van patiënten met aspecifieke nekpijn of lage-rugpijn bijvoorbeeld zijn associaties tussen de beweeglijkheid van de wervelkolom enerzijds en pijn en beperkingen in activiteiten anderzijds zwak of afwezig[36-41], hetgeen beoordelingen over consistentie in klinische informatie uiterst moeilijk maakt.

Studenten manuele therapie, evenals meer ervaren therapeuten, dienen voortdurend te worden geconfronteerd met de klinische onzekerheden rondom de subjectieve beoordelingen over bewegingsfunctiestoornissen, de werkingsmechanismen en effectiviteit van gewrichtsmobilisaties en –manipulaties, de complexe relaties tussen klinische variabelen en de waarde en interpretatie van patiëntgeoriënteerde uitkomsten. Het hedendaagse onderwijs op masterniveau stelt manueeltherapeuten in de gelegenheid te leren omgaan met deze onzekerheden door de kennis, vaardigheden en attitudes te ontwikkelen voor een geavanceerd niveau van *evidence-based* klinisch redeneren, de rechtvaardiging van hun besluitvorming, een kritische benadering van het praktisch handelen en een hoog niveau van zelfanalyse.[42] In Nederland is het naar een academisch niveau tillen van de manuele therapie via universitaire (vervolg)opleidingstrajecten nu een noodzakelijk te nemen volgende stap.

Sinds jaren bevinden wereldwijd nekpijn en lage-rugpijn zich samen met vele andere musculoskeletale aandoeningen onder de vooraanstaande oorzaken van beperkingen in het functioneren van mensen.[43] Patiënten met deze aandoeningen lijken helaas nog steeds nauwelijks beter af te zijn met de overdaad aan beschikbare behandelinterventies, inclusief manuele therapie aan de wervelkolom, ten opzichte van het te verwachten natuurlijke beloop van deze aandoeningen.[44;45] De manuele therapie bevindt zich nu in een kritieke fase waarin het zich kan bewijzen als een solide partner voor patiënten en de maatschappij zich baserend op nieuw, innovatief onderzoek, of die kans mislopen.

Ons onderzoek ondersteunt een multivariabele, biopsychosociale, hypothese-gestuurde benadering van het evalueren van de klinische diagnostiek en de klinische besluitvorming in de manuele therapie in tegenstelling tot het continueren van onderzoek naar de waarde van afzonderlijke tests.[3;4] Wij stellen methoden voor om een dergelijke benadering te incorporeren in betrouwbaarheidsonderzoek en gerandomiseerde klinische effectstudies met als uiteindelijk doel om die patiënten die reageren (of niet) op manuele therapie aan de wervelkolom beter te identificeren.[46]

Tot besluit keer ik terug naar de vraag die is ontstaan in het begin van mijn klinische carrière en die uiteindelijk de inspiratie vormde voor deze thesis: Was mijn onzekerheid bij het uitvoeren en interpreteren van het passief intervertebraal bewegingsonderzoek van de wervelkolom, in vergelijking met mijn zelfverzekerde docenten, stagebegeleiders en collega's, toe te schrijven aan mijn incompetentie of aan hun overmoedigheid? Ik twijfel er niet aan dat ik destijds nog onvoldoende vaardig was om als expert passieve segmentale intervertebrale beweeglijkheid te beoordelen, en hoogstwaarschijnlijk heb ik dat vereiste niveau nog steeds niet. Ik heb tenslotte mijn 10.000 uren van '*deliberate practice*' nog niet volgemaakt.[47] Anderzijds neigen ervaren manueeltherapeuten ernaar hun eigen prestaties te overschatten.[3;48] Gelukkig wordt een zekere mate van overmoedigheid beschouwd als een algemeen kenmerk van menselijk gedrag dat voordelig kan zijn op momenten dat definitieve beslissingen of acties zijn vereist.[49] Op zijn minst blijkt hieruit dat niets menselijks manueeltherapeuten vreemd is!

Referenties

1. Abbott JH, Flynn TW, Fritz JM, Hing WA, Reid D, Whitman JM. Manual physical assessment of spinal segmental motion: Intent and validity. Man Ther 2009;14:36-44.
2. MacDermid JC, Walton DM, Côté P, Santaguida PL, Gross A, Carlesso L, and ICON. Use of outcome measures in managing neck pain: An international multidisciplinary survey. Open Orthop J 2013;7:506-20.
3. Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine: A survey among physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51.
4. Van Trijffel E, Plochg T, Van Hartingsveld F, Lucas C, Oostendorp RAB. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual therapy. A qualitative interview study. J Man Manip Ther 2010;18:111-8.
5. Maher CG, Simmonds M, Adams R. Therapists' conceptualization and characterization of the clinical concept of spinal stiffness. Phys Ther 1998;78:289-300.
6. Maher C, Adams R. Is the clinical concept of spinal stiffness multidimensional? Phys Ther 1995;75:854-60.
7. Maher C, Adams R. A psychophysical evaluation of manual stiffness discrimination. Aust J Physiother 1995;41:161-7.
8. McGill S. Invited commentary on intrarater and interrater reliability of select clinical tests in patients referred for diagnostic facet joint blocks in the cervical spine. Arch Phys Med Rehabil 2013;94:1635-7.
9. Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69.
10. Bracht MA, Nunes GS, Celestino J, Schwertner DS, França LC, De Noronha M. Inter- and intra-observer agreement of the motion palpation test for lumbar vertebral rotational asymmetry. Physiother Canada 2015;67:169-73.
11. Brismée J-M, Atwood K, Fain M, Hodges J, Sperle A, Swaney M, Phelps V, Van Paridon D, Matthijs O, Sizer P. Interrater reliability of palpation of three-dimensional segmental motion of the lumbar spine. J Man Manip Ther 2005;13:215-20.

12. Brismée J-M, Gipson D, Ivie D, Lopez A, Moore M, Matthijs O, Phelps V, Sawyer S, Sizer P. Interrater reliability of a passive physiological intervertebral motion test in the mid-thoracic spine. J Manipulative Physiol Ther 2006;29:368-73.
13. Cleland JA, Childs JD, Fritz JM, Whitman JM. Interrater reliability of the history and physical examination in patients with mechanical neck pain. Arch Phys Med Rehabil 2006;87:1388-95.
14. Deore M, May S. The inter-rater and intra-rater reliability of passive physiological accessory movement assessment of lumbar spine in novice manual therapists. J Bodywork Movement Ther 2012;16:289-93.
15. Hanney WJ, George SZ, Kolber MJ, Young I, Salamh PA, Cleland JA. Inter-rater reliability of select physical examination procedures in patients with neck pain. Physiother Theory Pract 2014;30:345-52.
16. Johansson F. Interexaminer reliability of lumbar segmental mobility tests. Man Ther 2006;11:331-6.
17. Lundberg G, Gerdle B. The relationships between spinal sagittal configuration, joint mobility, general low back mobility and segmental mobility in female homecare personnel. Scand J Rehabil Med 1999;31:197-206.
18. Manning DM, Dedrick GS, Sizer PS, Brismée J-M. Reliability of a seated three-dimensional passive intervertebral motion test for mobility, end-feel, and pain provocation in patients with cervicalgia. J Man Manip Ther 2012;20:135-41.
19. Minaya Muñoz FJ, Valera Garrido F, Veiga Monasterioguren X. Fiabilidad interobservador del test de movilidad pasiva accesoria intervertebral lumbar y su correlación con la galvanopalpación en sujetos sanos. Rev Iberoam Fisioter Kinesiol 2010;13:4-9.
20. Ogince M, Hall T, Robinson K, Blackmore AM. The diagnostic validity of the cervical flexion-rotation test in C1/2-related cervicogenic headache. Man Ther 2007;12:256-62.
21. Piva SR, Erhard RE, Childs JD, Browder DA. Inter-tester reliability of passive intervertebral and active movements of the cervical spine. Man Ther 2006;11:321-30.
22. Schreiner M. Interrater-Reliabilität passiver physiologischer intervertebraler Bewegungen in der Sagittalebene der LWS. Manuelletherapie 2008;12:201-5.
23. Joshua AM, Celermajer DS, Stockler MR. Beauty is in the eye of the examiner: Reaching agreement about physical signs and their value. Intern Med J 2005;35:178-87.

24. Sturmberg JP, Martin CM, Katerndahl DA. Systems and complexity thinking in general practice literature: An integrative, historical narrative review. Ann Fam Med 2014;12:66-74.
25. Van de Pol RJ, Van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: A systematic review. J Physiother 2010;56:7-17.
26. Van Trijffel E, Van de Pol RJ, Oostendorp RAB, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review. J Physiother 2010;56:223-35.
27. Hazle CR Jr, Nitz AJ. A simulated passive intervertebral motion task: Observations of performance in a cross-sectional study. J Man Manip Ther 2012;20:121-9.
28. Haneline MT, Cooperstein R, Young M, Birkeland K. Spinal motion palpation: A comparison of studies that assessed intersegmental end feel vs excursion. J Manipulative Physiol Ther 2008;31:616-26.
29. Nyberg RE, Smith AR Jr. The science of spinal motion palpation: A review and update with implications for assessment and intervention. J Man Manip Ther 2013;21:160-7.
30. Davis DA. How to help professionals maintain and improve their knowledge and skills: Triangulating best practices in medicine. In: Ericsson KA (editor). Development of professional expertise. Toward measurement of expert performance and design of optimal learning environments. New York, NY: Cambridge University Press, 2009.
31. Ericsson KA. Enhancing the development of professional performance: Implications from the study of deliberate practice. In: Ericsson KA (editor). Development of professional expertise. Toward measurement of expert performance and design of optimal learning environments. New York, NY: Cambridge University Press, 2009.
32. Greenhalgh T, Howick J, Maskrey N, for the Evidence Based Medicine Renaissance Group. Evidence based medicine: A movement in crisis? BMJ 2014;348:g3725.
33. O'Dunn-Orto A, Hartling L, Campbell S, Oswald AE. Teaching musculoskeletal skills to medical trainees and physicians: A Best Evidence in Medical Education systematic review of strategies and their effectiveness: BEME Guide No. 18. Med Teach 2012;34:93-102.
34. Heemskerk MABM, Staal JB, Bierma-Zeinstra SMA, De Haan G, Hagenaars LHA, Lanser K, Van der Windt DAWM, Oostendorp RAB, Hendriks HJM. KNGF Guideline Complaints of Arms, Neck and or Shoulder, 2010. Available at: https://www.fysionet-

evidencebased.nl/images/pdfs/richtlijnen/kans_2010/kans_praktijkrichtlijn.pdf. [In Dutch] (accessed 24 November, 2014)

35. Mossman K. The complexity paradox. The more answers we find, the more questions we have. New York, NY: Oxford University Press, 2014.
36. Leboeuf-Yde C, Van Dijk J, Franz C, Hustad SA, Olsen D, Pihl T, Röbech R, Vendrup SS, Bendix T, Kyvik KO. Motion palpation findings and self-reported low back pain in a population-based study sample. J Manipul Physiol Ther 2002;25:80-7.
37. Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. Clin Rehabil 1999;13:211-8.
38. Olson SL, O'Connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. J Orthop Sports Phys Ther 2000;30:13-20.
39. Parks KA, Crichton KS, Goldford RJ, McGill SM. A comparison of lumbar range of motion and functional ability scores in patients with low back pain: Assessment for range of motion validity. Spine 2003;28:380-4.
40. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 1998;78:951-63.
41. Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. Phys Ther 2000;80:240-50.
42. Rushton A, Lindsay G. Defining the construct of masters level clinical practice in manipulative physiotherapy. Man Ther 2008;15:93-9.
43. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163-96.
44. Haldeman S, Dagenais S. A supermarket approach to the evidence-informed management of chronic low back pain. Spine J 2008;8-1-7.
45. Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: The HUNT study. Pain 2013;154:1237-44.
46. Van Trijffel E, Lindeboom R, Bossuyt PMM, Schmitt MA, Lucas C, Koes BW, Oostendorp R. Indicating spinal joint mobilisations or manipulations in patients with neck or low-back

pain: Protocol of an inter-examiner reliability study among manual therapists. Chiropr Man Therap 2014;22:22.
47. Ericsson KA. Expertise. Curr Biol 2014;24:R508-10.
48. Oostendorp RAB, Elvers H, Mikojalewska E, Laekeman M, Van Trijffel E, Samwel H, Duquet W. Manual physical therapists' use of biopsychosocial history taking in the management of patients with back or neck pain in clinical practice. ScientificWorldJournal 2015;2015:170463.
49. Croskerry P, Norman G. Overconfidence in clinical decision making. Am J Med 2008;121:S24-9.

Appendix

Onderzoeken gepubliceerd tussen 31 maart 2004 en 31 mei 2015*

Onderzoek (Eerste auteur)	**Beoordelingscriteria en schalen**	**Positie en beweging**	**Betrouwbaarheid**
Bracht[10]	Asymmetrie in rotatiebeweging (vergelijking links-rechts)	*In buikligging:* unilaterale P-A glij links en rechts (via processi transversi)	L2: κ 0,12 tot 0,23
Brismeé[11]	Richting van lateroflexie leidend tot de grootste rotatie (ipsilateraal of contralateraal)	*In linker zijligging:* rotatie rechts in links en rechts lateroflexie met de heupen en knieën respectievelijk in 45° en 70°	L4-L5: κ -0,08 tot 0,04
Brismeé[12]	Bewegingspatroon leidend tot de grootste rotatie (ipsilateraal of contralateraal)	*In zit:* rotatie in extensie en links en rechts lateroflexie	T6-T7: κ 0,27 tot 0,65
Cleland[13]	Normaal mobiel-hypermobiel-hypomobiel (vergelijking links-rechts of relatief ten opzichte van het segment boven en onder het geteste niveau)	*In rugligging:* A-P glij C0-C1 rotatie C1-C2 *In buikligging:* P-A glij C2-T9	C0-C1: κ_w -0,26, 0,46 C1-C2: κ_w 0,72, 0,74 C2 to C7: κ_w 0,01 tot 0,54 T1 to T9: κ_w 0,13 tot 0,82
Deore[14]	Hypomobiel-normaal-hypermobiel	*In buikligging:* P-A glij	L1: κ 0,20 L2: κ 0,19 L3: κ 0,03 L4: κ 0,00 L5: κ 0,01
Hanney[15]	Hypomobiel-normaal-hypermobiel	*In rugligging:* A-P glij C0-C1 rotatie C1-C2 laterale glij C2-T9	C0-C1: κ_w 0,15 C1-C2: κ_w 0,31 C2-C3: κ_w 0,30 C3-C4: κ_w 0,22 C4-C5: κ_w 0,43 C5-C6: κ_w 0,30 C6-C7: κ_w 0,23
Johansson[16]	Extreem hypomobiel-hypomobiel-normaal mobiel-hypermobiel-extreem hypermobiel	*In zijligging:* flexie, extensie met de heupen en knieën licht gebogen	L1-L2: κ_w -0,06 tot 0,27 L2-L3: κ_w -0,14 tot 0,54 L3-L4: κ_w -0,12 tot 0,28 L4-L5: κ_w 0,02 tot 0,17 L5-S1: κ_w -0,09 tot 0,56
Lundberg[17]	Extreem hypomobiel-middelmatig hypomobiel-normaal mobiel-middelmatig	*In zijligging:* algehele segmentale beweeglijkheid van flexie, extensie, rotatie links en	T10-T11: κ_w NG T11-T12: κ_w 0,73

	hypermobiel-extreem hypermobile	rechts en translatoire joint play met de heupen en knieën gebogen	T12-L1: κ_w 0,59 L1-L2: κ_w 0,68 L2-L3: κ_w 0,61 L3-L4: κ_w NG L4-L5: κ_w 0,75 L5-S1: κ_w 0,70
Manning[18]	Hypomobiliteit en hard, stug eindgevoel in vergelijking met het segment boven het geteste niveau	*In zit:* driedimensionale extensie en ipsilaterale lateroflexie en rotatie	C2-C3: *hypomobiliteit* κ 0,32, 0,33 en *eindgevoel* κ 0,28, 0,37 C3-C4: κ 0,33, 0,41 en κ 0,32, 0,50 C4-C5: κ 0,41, 0,48 en κ 0,39, 0,47 C5-C6: κ 0,21, 0,57 en κ 0,25, 0,60 C6-C7: κ 0,22, 0,58 en κ 0,28, 0,59
Minaya Muñoz[19]	Abnormaal eindgevoel	*In buikligging:* unilaterale P-A glij links en rechts	L1-L2: PO *rechts* 86,6%, *links* 86,6% L2-L3: PO 83,3%, 83,3% L3-L4: PO 90,0%, 90,0% L4-L5: PO 86,6%, 93,3% L5-S1: PO 86,6%, 76,6% Overall κ 0,50
Ogince[20]	Beperkte bewegingsuitslag gebaseerd op de interpretatie van de therapeut van een vermindering met 10° en op een hard eindgevoel	*In rugligging:* links en rechts rotatie met de cervicale wervelkolom in volledige flexie	C1-C2: κ 0,81
Piva[21]	Normaal mobiel-hypomobiel	*In rugligging:* laterale glij C0-C1 laterale verplaatsing C1 links en rechts rotatie met de cervicale wervelkolom in volledige flexie laterale glij C2-C6	C0-C1: κ 0,81 C1: κ 0,35 C1-C2: κ 0,30 C2: κ 0,46 C3: κ 0,25 C4: κ 0,27 C5: κ 0,18 C6: κ -0,07
Schreiner[22]	Normaal mobiel-hypermobiel-hypomobiel	*In zijligging:* flexie, extensie met de heupen en knieën gebogen	T12-L1: κ *extensie* -0,11, *flexie* κ NG L1-L2: κ NG, κ NG L2-L3: κ 0,00, κ -0,07 L3-L4: κ 0,29, κ -0,04 L4-L5: κ -0,24, κ -0,14 L5-S1: κ 0,02, κ -0,02

A-P: antero-posterior, κ: kappa, κ_w: gewogen kappa, NG: niet gerapporteerd, P-A: postero-anterior, PO: percentage overeenstemming,

*Uitgezonderd Lundberg: gepubliceerd vóór 31 maart 2004

PhD Portfolio

PhD Portfolio

Name: Emiel van Trijffel

PhD period: 2006-2015

Supervisors: Prof. Dr. P.M.M. Bossuyt
Prof. Dr. R.A.B. Oostendorp
Prof. Dr. C. Lucas

1. PhD training

	Year	Workload (ECTS)
General courses		
Master of Science Education in Clinical Epidemiology**,** University of Amsterdam (Academic Medical Centre)		
Introduction to Science/Basic Epidemiology	2002	10
Advanced Epidemiology	2003	4
Advanced Biostatistics	2003	6
Health Care Policy	2003	6
Clinical Uncertainty	2003	5
Advanced Epidemiology/Clinimetrics	2003	6
Health Economics	2003	5
Evidence-Based Practice/Systematic Reviews and Clinical Guidelines	2004	13
Capita Selecta	2004	5
Masterthesis	2005	24
		Total: 84
Specific courses		
Qualitative Research, University of Amsterdam (Academic Medical Centre)	2007	5
Oral presentations - National		
'Epidemiologische concepten'. SOMT Congres 'Cervicale manipulaties: Risk or benefit?', Amersfoort, the Netherlands	2013	0.5
'De rol en positie van passief segmentaal bewegingsonderzoek binnen het klinisch redeneren'. NVMT Congres 'Spinal management', Veldhoven, the Netherlands	2008	0.5
'Gebruik, interpretatie en percepties van passief segmentaal bewegingsonderzoek'. KNGF Jaarcongres Fysiotherapie 'Fysiotherapie: een sterk merk', Amsterdam, the Netherlands	2007	0.5

'Evidence-Based diagnostiek in de manuele therapie'. 1e Lustrumsymposium Universitaire Master Evidence-Based Practice, Universiteit van Amsterdam (Academic Medical Centre), Amsterdam, the Netherlands	2007	0.5
'Tussenbeoordelaarsbetrouwbaarheid van passief onderzoek naar intervertebrale beweeglijkheid van de cervicale en lumbale wervelkolom'. KNGF Jaarcongres Fysiotherapie 'Specialiseren doe je samen', the Hague, the Netherlands	2005	0.5
Poster 'Tussenbeoordelaarsbetrouwbaarheid van passief onderzoek naar intervertebrale beweeglijkheid van de cervicale en lumbale wervelkolom: een systematische review'. NVMT Congres 'Evidence Based Practice op de werkvloer', Veldhoven, the Netherlands	2005	0.5
Oral presentations - International		
'Connecting science to clinical practice: A mixed methods approach for exploring passive intervertebral motion assessment in manual therapy'. International IFOMT congress 'Connecting science to quality of life', Rotterdam, the Netherlands	2009	0.5
Other		
Research Group 'Tailored diagnostics and care in comprehensive care management: physical therapy aspects', University of Applied Sciences, Amsterdam School of Health Professions, Amsterdam, the Netherlands	2008-2009	8

2. Teaching

	Year	Workload (ECTS)
Lecturing, tutoring, mentoring, developing		
Master of Science Clinical Epidemiology, University of Amsterdam (Academic Medical Centre), Amsterdam, the Netherlands	From 2006	184
Master of Science Manual Medicine, Vrije Universiteit Brussel, Brussels, Belgium/Institute for Master Education in Musculoskeletal Therapy (SOMT), Amersfoort, the Netherlands	From 2013	6
Professional Masters Musculoskeletal Therapy, Institute for Master Education in Musculoskeletal Therapy (SOMT), Amersfoort, the Netherlands	From 2008	282
Minor Science, Amsterdam School of Health Professions, Amsterdam, the Netherlands	2008-2009	2
Supervising		
Master of Science Clinical Epidemiology, University of Amsterdam (Academic Medical Centre), Amsterdam, the Netherlands (21 theses)	From 2006	7.5

Professional Masters Musculoskeletal Therapy, Institute for Master Education in Musculoskeletal Therapy (SOMT), Amersfoort, the Netherlands (49 theses)	From 2008	7

Other		
Program Board Master of Science Manual Therapy, Vrije Universiteit Brussel, Brussels, Belgium	From 2013	1
Visitation panel Accreditation Organisation of the Netherlands and Flanders (NVAO)	2014-2015	1
Dutch Educational network Masters of Physiotherapy (DEMP)	From 2012	1.5

3. Publications

International peer reviewed	**Year**
De Roos P, **Van Trijffel E**, Strijbos JH, Lucas C. Effectiveness of a combined exercise training and home-based walking program on physical activity compared to usual care in moderate COPD. A randomized controlled trial. *Submitted*	2015
Oostendorp RAB, Elvers H, Mikojalewska E, Roussel N, **Van Trijffel E**, Samwel H, Nijs J, Duquet W. Cervicocephalagiaphobia: A subtype of phobia in patients with cervicogenic headache and neck pain? A pilot study. *Accepted*	2015
Oostendorp RAB, Elvers H, Mikojalewska E, Laekeman M, **Van Trijffel E**, Samwel H, Duquet W. Manual physical therapists' use of biopsychosocial history taking in the management of patients with back or neck pain in clinical practice. ScientificWorldJournal 2015;2015:170463	2015
Stenneberg MS, Schmitt MA, **Van Trijffel E**, Schröder CD, Lindeboom R. Validation of a new questionnaire to assess the impact of Whiplash Associated Disorders: The Whiplash Activity and participation List (WAL). Man Ther 2015;20:84-9	2015
Schrama PPM, Stenneberg MS, Lucas C, **Van Trijffel E**. Intra-examiner reliability of hand-held dynometry in the upper extremity. A systematic review. Arch Phys Med Rehabil 2014;95:2444-69	2014
Koehorst MLS, **Van Trijffel E**, Lindeboom R. Evaluative measurement properties of the Patient Specific Functional Scale for primary shoulder complaints in physical therapy practice. J Orthop Sports Phys Ther 2014;44:595-603	2014
Van Trijffel E, Lindeboom R, Bossuyt PMM, Schmitt MA, Lucas C, Koes BW, Oostendorp R. Indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain: Protocol of an inter-examiner reliability study among manual therapists. Chiropr Man Therap 2014;22:22	2014
Kappetijn O, **Van Trijffel E**, Lucas C. Efficacy of passive extension mobilization in addition to exercise in the osteoarthritic knee: An observational parallel-group study. Knee 2014;21:703-9	2014
Heus P, **Van Trijffel E**, Busch EM, Lucas C. Mydriatic visual acuity in diabetic patients: A randomized controlled trial. Optom Vis Sci 2013;90:249-56	2013
Van Trijffel E, Van de Pol RJ, Oostendorp RAB, Lucas C. Inter-rater reliability for measurement of passive physiological movements in lower extremity joints is generally low: A systematic review. J Physiother 2010;56:223-35	2010

Van der Sleen MI, Slot DE, **Van Trijffel E**, Winkel EG, Van der Weijden GA. Effectiveness of mechanical tongue cleaning on breath odeur and tongue coating: a systematic review. Int J Dent Hyg 2010;8:258-68 2010

Van Trijffel E, Plochg T, Van Hartingsveld F, Lucas C, Oostendorp RAB. The role and position of passive intervertebral motion assessment within clinical reasoning and decision-making in manual therapy. A qualitative interview study. J Man Manip Ther 2010;18:111-8 2010

Van de Pol RJ, **Van Trijffel E**, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: A systematic review. J Physiother 2010;56:7-17 2010

Van Trijffel E, Oostendorp RAB, Lindeboom R, Bossuyt PMM, Lucas C. Perceptions and use of passive intervertebral motion assessment of the spine: A survey among physiotherapists specializing in manual therapy. Man Ther 2009;14:243-51 2009

Bakker EWP, Verhagen AP, **Van Trijffel E**, Lucas C, Koes BW. Spinal mechanical load as a risk factor for low back pain. A systematic review of prospective cohort studies. Spine 2009;34:E281-93 2009

Bakker EWP, Verhagen AP, **Van Trijffel E**, Lucas C, Koning HJ, Koes BW. Individual advice in addition to standard guideline care in patients with acute non-specific low back pain. A survey on feasibility among physiotherapists and patients. Man Ther 2009;14:68-74 2009

Van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. Man Ther 2005;10:256-69 2005

Letters to the Editor

Scholten-Peeters GG, **Van Trijffel E**, Hutting N, Castien RF, Rooker S, Verhagen AP. Risk reduction of serious complications from manual therapy: Are we reducing the risk? Correspondence to: Rushton A, Rivett D, Carlesso L, Flynn T, Hing W, Kerry R. International framework for examination of the cervical region for potential of Cervical Arterial Dysfunction prior to Orthopaedic Manual Therapy intervention. Man Ther 2014;19:e5-6 2014

Bakker EW, Verhagen AP, **Van Trijffel E**, Lucas C. Koes BW. Response to: Olsen O. Re: Bakker EW, Verhagen AP, Van Trijffel E, Lucas C. Koes BW. Spinal mechanical load as a risk factor for low back pain: a systematic review of prospective cohort studies. Spine 2010;35:E576-7 2010

National

Scholten-Peeters W, Rooker S, **Van Trijffel E**, Verhagen A. De fysiotherapeutische diagnose. Toch ook een plaats voor de ICD? Reactie op: Van Enck, Nieuwenhuijzen. De fysiotherapeutische diagnose gediagnosticeerd. FysioPraxis februari 2014:22-3 2014

Van Trijffel E. Hoe betrouwbaar is passief segmentaal bewegingsonderzoek van de wervelkolom? Stimulus 2005;24:470-87 2005

Other

Reviewer for international journals From 2004

Dankwoord

Dankwoord

Hooggeleerde Bossuyt, beste Patrick, met veel bewondering heb ik geluisterd en gekeken naar de overstijgende blik waarmee jij vraagstukken benadert en waardoor ik veel heb geleerd over mijn eigen vakgebied, de manuele therapie. Daarnaast ben je een voortreffelijke coach die op de juiste momenten de sturing gaf die nodig was.

Hooggeleerde Oostendorp, beste Rob, het emeritaat was voor jou nooit aanleiding om je inzet en betrokkenheid bij dit promotietraject ook maar enigszins te laten vieren. Van jou heb ik geleerd dat het best mag en ook heel noodzakelijk is om geregeld aan de grondvesten van je eigen beroep te schudden. Jij was de eerste gepromoveerde manueeltherapeut in Nederland; het is voor mij een grote eer om jouw laatste promovendus te mogen zijn geweest.

Hooggeleerde Lucas, beste Cees, iedere promovendus heeft iemand nodig die hem het vertrouwen en de motivatie geeft om een promotietraject te starten én om ermee door te gaan. Ik prijs mij gelukkig dat jij diegene voor mij was, ook toen we niet meer direct samenwerkten in het AMC.

Hooggeachte leden van de Promotiecommissie, ik dank u bijzonder hartelijk voor uw bereidheid dit proefschrift te beoordelen en plaats te nemen in de oppositie.

Sander en Lotte, in het AMC en bij SOMT mijn sparringpartners en steunpilaren. Wat een eer en een fijn gevoel om jullie tijdens de verdediging naast mij te weten.

Mijn medeauteurs, die zonder uitzondering de hoofdstukken naar een hoger niveau hebben getild. In het bijzonder Prof. Dr. Cattrysse en Prof. Dr. Scafoglieri, beste Erik en Aldo, jullie verrijkten mij met jullie wereld van het fundamentele onderzoek in de manuele therapie, en Dr. Lindeboom, beste Robert, bij jou kon ik altijd terecht om mijn methodologische en statistische problemen met behulp van jouw briljante ingevingen op te lossen.

Drs. Smeets, beste Willy, bedankt voor de tijd en ruimte die je me gaf in een periode dat ik die nodig had.

Lieve schoonouders, Martin, Mischa & Tomas, familie, vrienden en alle collega's uit het AMC en van SOMT, veel dank voor jullie nimmer aflatende belangstelling en aanmoedigingen.

Mijn lieve ouders, jullie zijn er altijd voor ons met raad en daad.

Allerliefste Sylvia, mijn engel, je staat als laatste in dit Dankwoord maar helemaal bovenaan in mijn leven.

Curriculum Vitae

Curriculum Vitae

Emiel van Trijffel werd geboren op 19 maart 1968 te Rotterdam. Na het behalen van zijn vwo-diploma aan de Libanon Scholengemeenschap te Rotterdam (1986) studeerde hij fysiotherapie aan de Hogeschool Rotterdam & Omstreken (1986-1990). In 1992 startte hij zijn zelfstandige praktijk voor fysiotherapie in Rotterdam-Kralingen. Zijn belangstelling voor de methodologie en statistiek werd gewekt tijdens en na het in 1998 afronden van de opleiding manuele therapie aan de Stichting Opleiding Manuele Therapie (SOMT, thans: Stichting Opleidingen Musculoskeletale Therapie). Dit resulteerde in het volgen van de cursussen Scholing in Wetenschap I en II van het Nederlands Paramedisch Instituut (2001-2002) o.l.v. Prof. Dr. Rob Oostendorp en drs. Hans Elvers gevolgd door de universitaire masteropleiding Evidence Based Practice van de Universiteit van Amsterdam (Academisch Medisch Centrum, Hoofd thans: Prof. Dr. Cees Lucas). Na het cum laude behalen van zijn diploma in 2005 verbond hij zich aan deze laatste opleiding als coördinator en docent in de vakken Evidence-Based Practice, klinische epidemiologie en biostatistiek. Deze verbintenis duurde tot februari 2012. Momenteel begeleidt hij klinisch epidemiologen i.o. van deze master bij hun afstudeeronderzoeken. In 2008 had hij inmiddels zijn diepgekoesterde fysiotherapiepraktijk beëindigd en een functie als hoofd van de HBO-masteropleiding Manuele Therapie bij SOMT aanvaard. Sinds 2012 bekleedt hij bij deze instelling fulltime de functie van Hoofd Onderwijs, waaronder naast de master Manuele Therapie tevens de masters Bekkenfysiotherapie, Sportfysiotherapie, Fysiotherapie in de Geriatrie, de Master of Science Manuele Geneeskunde (i.s.m. de Vrije Universiteit Brussel en de Stichting Manuele Geneeskunde) en de afdeling Lifelong Learning vallen.