



STUDY PROTOCOL

Open Access

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

Emiel van Trijffel^{1,2*}, Robert Lindeboom¹, Patrick MM Bossuyt¹, Maarten A Schmitt², Cees Lucas¹, Bart W Koes³ and Rob AB Oostendorp^{4,5}

Abstract

Background: Manual spinal joint mobilisations and manipulations are widely used treatments in patients with neck and 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 daily 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, alone 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 toward 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 clinical trials. Potential methodological limitations of this study are discussed.

Keywords: Manual therapy, Motion assessment, Diagnostics, Decision-making, Reliability, Clinical reasoning, Neck pain, Back pain

* Correspondence: e.vantrijffel@somt.nl

¹Department of Clinical Epidemiology, Biostatistics & Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands

²Institute for Master Education in Musculoskeletal Therapy, Amersfoort, the Netherlands

Full list of author information is available at the end of the article

Background

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 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 and 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, alone 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 clavicle, 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 daily 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 (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 and 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 generalized chance-corrected reliability across all pairs of therapists. See Additional file 1 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, all 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 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 daily 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 favorably 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

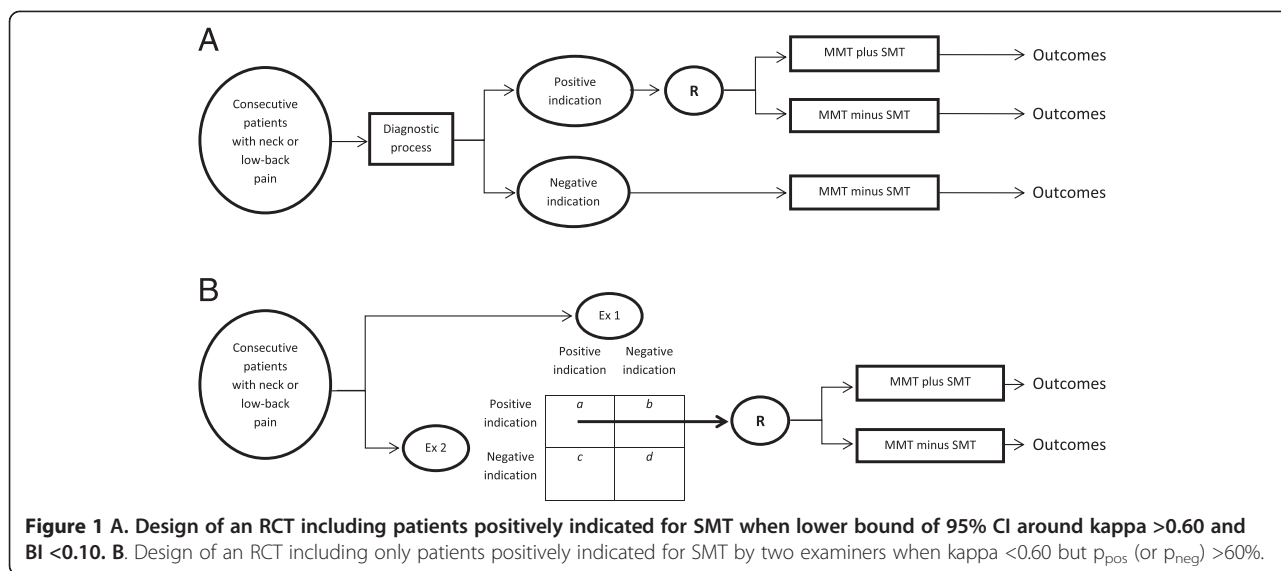
patterns [84]. With the aim to identify patients that may be more likely to show a positive response to spinal manipulation, clinical prediction rules have been derived to identify predictors in patients with neck and low-back pain [44-47,50-55]. Unfortunately, systematic reviews have consistently concluded that there is, as yet, insufficient evidence to support the general application of these rules [85-89]. Another systematic review found significant treatment effects favoring 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 clinical 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 daily practice, may outweigh the disadvantages of the current selection procedures in randomised clinical trials.

Awaiting evidence from the further validation of prediction rules and other classification systems, our study could offer an initial step toward 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 clinical 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 1A). 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 1B). 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 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 daily 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 posteroanterior 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 and 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 purpose 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.

Additional file

Additional file 1: Formulas for kappa and associated measures.

Abbreviations

ICF: International Classification of Functioning, Disability and Health;
PIVM: Passive intervertebral motion; SMT: Spinal manual therapy.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EVt is the principal investigator of the study, developed the research questions and methods, obtained ethical approval, and drafted the article. RL, MS, CL, BK, RO assisted in the development of the methods and wrote the study protocol. RL, EVt, PB developed the statistical plan for this protocol. PB, RO supervised the project. All authors assisted with revisions to the study protocol and methods, and approved the final version of the article.

Author details

¹Department of Clinical Epidemiology, Biostatistics & Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands. ²Institute for Master Education in Musculoskeletal Therapy, Amersfoort, the Netherlands. ³Department of General Practice, Erasmus MC University Medical Centre, Rotterdam, the Netherlands. ⁴Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands. ⁵Department of Rehabilitation, Physiotherapy and Manual Therapy, Faculty of Medicine and Pharmacology, Free University of Brussels, Brussels, Belgium.

Received: 25 January 2014 Accepted: 14 May 2014

Published: 20 June 2014

References

1. Borghouts JAJ, Koes BW, Bouter LM: Cost-of-illness in neck pain in the Netherlands in 1996. *Pain* 1999, **80**:629-636.
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**:1689-1698.
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**(Suppl 4):39-51.
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**:1457-1463.
5. van Tulder MW, Koes BW, Bouter LM: A cost-of-illness study of back pain in The Netherlands. *Pain* 1995, **62**:233-240.
6. Waddell G: Low back pain: a twentieth century health care enigma. *Spine* 1996, **21**:2820-2825.
7. Assendelft WJ, Morton SC, Yu EI, Suttrop MJ, Shekelle PG: Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 2003, **138**:871-881.
8. Gross A, Miller J, D'Sylva J, Burnie SJ, Goldsmith CH, Graham N, Haines T, Branfort G, Hoving JL: Manipulation or mobilization for neck pain: a Cochrane Review. *Man Ther* 2010, **15**:315-333.
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**:531-538.
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, Branfort G, Hoving JL: Manipulation or mobilisation for neck pain. *Cochrane Database Syst Rev* 2010, (Issue 1): Art. No.: CD004249. doi:10.1002/14651858.CD004249.pub3.
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**:655-662.
13. Rubinstein SM, van Middelkoop M, Assendelft WJJ, de Boer MR, van Tulder MW: Spinal manipulative therapy for chronic low-back pain. *Cochrane Database Syst Rev* 2011, (Issue 2): Art. No.: CD008112. doi:10.1002/14651858.CD008112.pub2.
14. Rubinstein SM, Terwee CB, Assendelft WJJ, de Boer MR, van Tulder MW: Spinal manipulative therapy for acute low-back pain. *Cochrane Database Syst Rev* 2012, (Issue 9): Art. No.: CD008880. doi:10.1002/14651858.CD008880.pub2.
15. Farrell JP, Jensen GM: Manual therapy: A critical assessment of role in the profession of physical therapy. *Phys Ther* 1992, **72**:843-852.
16. Maher C, Latimer J: Pain or resistance - the manual therapists' dilemma. *Austr J Physiotherapy* 1992, **38**:257-260.

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**:113–124.
18. Jull G, Treleaven J, Versace G: **Manual examination: Is pain provocation a major cue for spinal dysfunction?** *Austr J Physiotherapy* 1994, **40**:159–165.
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**:243–251.
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**:440–460.
22. Bartko JJ, Carpenter WT: **On the methods and theory of reliability.** *J Nerv Ment Dis* 1976, **163**:307–317.
23. de Vet HC, Terwee CB, Knol DL, Bouter LM: **When to use agreement versus reliability measures.** *J Clin Epidemiol* 2006, **59**:1033–1039.
24. Streiner DL, Norman GR: *Health measurement scales. A practical guide to their development and use.* 4th edition. Oxford, UK: Oxford University Press; 2008:167–210.
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**:616–626.
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–E425.
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**:475–485.
30. van Trijffel E, Anderegg O, 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**:256–269.
31. Moons KG, Biesheuvel CJ, Grobbee DE: **Test research versus diagnostic research.** *Clin Chem* 2004, **50**:473–476.
32. Jones MA, Jensen G, Edwards I: **Clinical reasoning in physiotherapy.** In *Clinical Reasoning in the Health Professions*. 3rd edition. Edited by Higgs J, Jones MA, Loftus S, Christensen N. Edinburgh, UK: Elsevier/Butterworth Heinemann; 2008:245–256.
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**:455–470.
34. Rivett DA, Higgs J: **Hypothesis generation in the clinical reasoning behaviour of manual therapists.** *J Phys Ther Educ* 1997, **11**:40–45.
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**:111–118.
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**:167–175.
37. Riddle DL: **Classification and low back pain: A review of the literature and critical analysis of selected systems.** *Phys Ther* 1996, **78**:708–737.
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**(Suppl 2):19–42.
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**:213–221.
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**:42–48.
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**:72–79.
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**:920–928.
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**:2835–2843.
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**:173–190.
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**:1753–1762.
48. Cook C: **Key issues for manual therapy clinical practice and research in North America.** *Man Ther* 2013, **18**:269–270.
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**:513–524.
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**:1619–1632.
52. Puentedura EJ, Cleland JA, Landers MR, Mintken PE, Louw A, Fernández-de-las-Peñas C: **Development of a clinical prediction rule to identify patients with neck pain likely to benefit from thrust joint manipulation to the cervical spine.** *J Orthop Sports Phys Ther* 2012, **42**:577–592.
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**:670–680.
54. Thiel HW, Bolton JE: **Predictors for immediate and global responses to chiropractic manipulation of the cervical spine.** *J Manipulative Physiol Ther* 2008, **31**:172–183.
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**:306–315.
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**:50–55.
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**(Suppl 4):14–23.
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**(Suppl 2):192–300.
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**(Suppl 11):240–252.

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**:268–275.
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; 2006:5–48.
64. **KNGF Guideline Low-back Pain.** 2013. [<http://www.fysionet-evidencebased.nl/index.php/component/kngf/richtlijnen/lage-rugpijn-2013>] [In Dutch] Last accessed 30 December 2013.
65. van der El A: *Orthopaedic manual therapy diagnosis. Spine and temporomandibular joints.* London, UK: Jones and Bartlett Publishers; 2010:351–498.
66. Macfadyen N, Maher CG, Adams R: **Number of sampling movements and manual stiffness judgements.** *J Manipulative Physiol Ther* 1998, **21**:604–610.
67. **International classification of functioning, disability and health (ICF).** [<http://www.who.int/classifications/icf/en/>] Last accessed 30 December 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**:551–558.
70. Byrt T, Bishop J, Carlin JB: **Bias, prevalence and Kappa.** *J Clin Epidemiol* 1993, **46**:423–429.
71. Feinstein AR, Cicchetti DV: **High agreement but low kappa: I. The problems of two paradoxes.** *J Clin Epidemiol* 1990, **43**:543–549.
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**:127–134.
73. Landis JR, Koch DG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33**:159–164.
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**:2880–2884.
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**:1992–1996.
78. Bouter LM, van Tulder MW, Koes BW: **Methodologic issues in low back pain research in primary care.** *Spine* 1998, **23**:2014–2020.
79. **Clinical research agenda for physical therapy.** *Phys Ther* 2000, **80**:499–513. <http://www.ncbi.nlm.nih.gov/pubmed/10792860>.
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**:165–174.
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**:1022–1031.
84. Kent P, Keating J: **Classification in nonspecific low back pain: What methods do primary care clinicians currently use?** *Spine* 2005, **30**:1433–1440.
85. Beniuk JM, Bishop MD, George SZ: **Clinical prediction rules for physical therapy interventions: A systematic review.** *Phys Ther* 2009, **89**:114–124.
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, Mjosund 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**:843–854.
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**:762–769.
90. Slater SL, Ford JJ, Richards MC, Taylor NF, Surrick LD, Hahne AJ: **The effectiveness of sub-group specific manual therapy for low back pain: A systematic review.** *Man Ther* 2012, **17**:201–212.
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**:384–390.
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**:460–468.
93. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *JAMA* 1999, **282**:1061–1066.
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**:469–476.
96. Scales CD Jr, Canfield SE: **Advanced topics in evidence-based urological oncology: Using results of a subgroup analysis.** *Urol Oncol* 2011, **29**:462–466.
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**:200–205.
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**:471–485.
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**:108–117.
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**:438–444.
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**:219–226.
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**:718–724.
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**:484–495.
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**:406–411.
108. Jette AM, Delitto A: **Physical therapy treatment choices for musculoskeletal impairments.** *Phys Ther* 1997, **77**:145–154.
109. Li LC, Bombardier C: **Physical therapy management of low back pain: An exploratory survey of therapist approaches.** *Phys Ther* 2001, **81**:1018–1028.
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**:189–196.
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**:1151–1167.
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**:1040–1051.

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**:206–227.
115. Jull G: Use of high and low velocity cervical manipulative therapy procedures by Australian manipulative physiotherapists. *Aust J Physiother* 2002, **48**:189–193.
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**:10–18.
117. Sim J, Wright CC: The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005, **85**:257–268.
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**:106–110.
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**:1634–1641.
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**:1313–1318.
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**:191–198.

doi:10.1186/2045-709X-22-22

Cite this article as: van Trijffel *et al.*: Indicating spinal joint mobilisations or manipulations in patients with neck or low-back pain: protocol of an inter-examiner reliability study among manual therapists. *Chiropractic & Manual Therapies* 2014 **22**:22.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

