Clinical Scenarios for which Cervical Mobilization and Manipulation Are Considered by an Expert Panel to be Appropriate (and Inappropriate) for Patients with Chronic Neck Pain

Patricia M Herman, ND, PhD,

RAND Corporation, 1776 Main Street, PO Box 2138, Santa Monica, CA 90403

Howard Vernon, DC, PhD,

Division of Research, Canadian Memorial Chiropractic College

Eric L Hurwitz, DC, PhD,

Office of Public Health Studies, University of Hawai`i at M noa

Paul G Shekelle, MD, PhD,

RAND Corporation

Margaret D Whitley, MPH,

RAND Corporation

Ian D Coulter, PhD

RAND Corporation

Abstract

Objectives.—Cervical mobilization and manipulation are two therapies commonly used for chronic neck pain (CNP). However, the safety, especially of cervical manipulation, is controversial. This study identifies the clinical scenarios for which an expert panel rated cervical mobilization and manipulation as appropriate and inappropriate.

Methods.—An expert panel, following a well-validated modified-Delphi approach, used an evidence synthesis and clinical acumen to develop and then rate the appropriateness of cervical mobilization and manipulation for each of an exhaustive list of clinical scenarios for CNP. Key patient characteristics were identified using decision tree analysis (DTA).

Results.—Three hundred seventy-two clinical scenarios were defined and rated by an 11-member expert panel as to the appropriateness of cervical mobilization and manipulation. Across clinical scenarios more were rated inappropriate than appropriate for both therapies, and more scenarios were rated inappropriate for manipulation than mobilization. However, the number of patients presenting with each scenario is not yet known. Nevertheless, DTA indicates that all

(corresponding author) Ph: 310-393-0411 x7129; Fax: 310-260-8161; pherman@rand.org.

clinical scenarios that included red flags (e.g., fever, cancer, inflammatory arthritides or vasculitides), and some others involving major neurologic findings, especially if previous manual therapy was unfavorable, were rated as inappropriate for both cervical mobilization and manipulation. DTA also identified the absence of cervical disc herniation, stenosis, or foraminal osteophytosis on additional testing as the most important patient characteristic in predicting ratings of appropriate.

Conclusions.—Clinical guidelines for CNP should include information on the clinical scenarios for which cervical mobilization and manipulation were found inappropriate, including those with red flags, and others involving major neurologic findings if previous manual therapy was unfavorable.

Keywords

chronic neck pain; appropriateness of care; RAND/UCLA Appropriateness Method; decision tree analysis; cervical mobilization and manipulation

Introduction

All patients should receive appropriate care, but challenges lie in defining what constitutes appropriate care, and in identifying the therapies that would be appropriate for each patient. In response to these challenges the RAND Corporation and the University of California, Los Angeles, (UCLA) developed a method to study the appropriateness of care.[1-5] The RAND/UCLA Appropriateness Method (RUAM) makes it feasible to take the best of what is known from research and apply it—using the knowledge of experienced clinicians—over the wide range of patients and health problems seen in real-world clinical practice. Clinicians are, after all, the final translators of evidence into practice, and this approach formalizes the process.

The RUAM is an expert panel-based, modified-Delphi approach that has been widely used and studied and found to be both reliable and valid. The test-retest reliability for the same panelists 6–8 months later was >.90,[6] and results across several panels for the same procedure were acceptably reproducible (kappa statistics 0.5 to 0.7).[7, 8] The RUAM estimates have also been found to be consistent with the literature and follow a logical clinical rationale,[6] and to have favorable predictive ability—i.e., patients receiving care rated as appropriate have been found to have better outcomes,[9-13] and later clinical trials targeting specific clinical scenarios have validated panelists' ratings.[11]

Chronic neck (cervical) pain (CNP) is a common type of chronic pain,[14, 15] and cervical mobilization and manipulation are two therapies that have been used to treat CNP.[16-18] However, there is long-standing controversy over their safety. Several studies have shown associations between cervical manipulation and neurovascular complications, including vertebral arterial dissection and stroke.[19-23] Others have shown an almost equal association between a visit to a general practitioner and stroke, and posit that the early symptoms of a pending stroke are what precipitate both types of visits.[24, 25] Others argue that any movement of the neck, and not specifically manipulation, is a potential trigger for someone at risk for a stroke.[26] Given this controversy, that back and neck pain are the

most common indications for receiving spinal manipulation,[17] and that adverse events might be preventable with good clinical judgement,[27] determining the appropriateness (the amount by which benefits exceed risks) of cervical mobilization and manipulation for patients with different CNP presentations is a worthy objective.

This study used the RUAM to determine the appropriateness of cervical mobilization and manipulation for patients with different presentations of CNP, and the key patient characteristics associated with appropriateness. A separate paper will apply these results to determine the prevalence of appropriate and inappropriate care for CNP in chiropractic practice.

Methods

The RUAM[3, 28, 29] utilizes a panel of clinical and content experts to translate the available evidence on a therapy into ratings of the level of its appropriateness for each type of patient (clinical scenario) who might present with the condition of interest (here, CNP). Panelists were chosen based on their clinical expertise across the different disciplines and specialties who treat CNP, and diversity of geographic location. They were identified through their publications, professional reputation, and from our content experts, and some served on an earlier acute neck pain panel.[30, 31] The 11-member panel included one orthopedist, one osteopath, one internist, five chiropractors, one neurologist (who is also a chiropractor), one physical therapist, and one physiatrist. Each panelist received a $1000 honorarium.

Panelists were first presented with a detailed systematic review of latest evidence on the effectiveness and safety of cervical mobilization and manipulation for CNP,[18] and then asked to rate on a 1–9 scale the extent to which the benefits of each therapy outweigh its risks for each clinical scenario. Ratings of 7–9 (i.e., the therapy is appropriate) were given if: "The expected health benefit (e.g., increased life expectancy, relief of pain, reduction in anxiety, improved functional capacity) exceeds the expected negative consequences (e.g., mortality, morbidity, anxiety, pain, time lost from work) by a sufficiently wide margin that the procedure is worth doing, exclusive of cost."[2] The instructions given to panelists and definitions of terms used in the rating process are found in a detailed, publicly available RAND report.[32]

The clinical scenarios to rate were organized into sections for ease of rating—i.e., once one (the first) clinical scenario in a section was rated, the others only differed by one or two patient characteristics and could be evaluated quickly.[32] Clinical scenarios categorized patients using the usual criteria a physician would use to choose a therapy: history, symptoms, physical and radiographic findings, and response to prior treatment. The list of scenarios to rate was designed to be comprehensive (capture all types of patients with CNP), detailed (the procedure should be equally appropriate or inappropriate for all patients in a scenario), and manageable (all scenarios could be rated within a reasonable amount of time). Scenarios deemed implausible were dropped. A panelist who is used to the process can rate 150–200 scenarios an hour on average.[3] The list of clinical scenarios was based on the literature review, clinical expert advice, and the list of scenarios used for an earlier study on

cervical manipulation for acute neck pain.[30, 31] The initial at-home round of the RUAM was applied to 386 clinical scenarios each for cervical mobilization and manipulation. In each case, 193 clinical scenarios were rated under two conditions: (1) there has been no other adequate conservative care for this episode or (2) non-manual conservative care for this episode had failed.

Panelists rated each clinical scenario twice: 1) first individually at home; and 2) then during a one-day face-to-face meeting and after discussion with the other panelists. At the beginning of the face-to-face meeting each panelist was given a personalized printout that identified his or her at-home ratings in relation to the distribution, but not the identities, of all other panelists' ratings. The home and face-to-face rating sessions took place in April and May of 2015, respectively. Panelists were asked to make their ratings using their own best clinical judgement and content knowledge (rather than their perceptions of what other experts might say) and considering an average patient currently presenting to an average North American practitioner who performs this procedure in an average care-providing facility. Consensus was reported, but not required.

For this study, we defined manipulation of the cervical spine as "a controlled, judiciously applied dynamic thrust (adjustment), that may include extension and rotation of the cervical region, of high or low velocity and low amplitude force directed to spinal joint segment within patient tolerance."[32]p20 Mobilization of the cervical spine was defined as "a controlled, judiciously applied force of low velocity and variable amplitude directed to spinal joint segments. These procedures usually do not take joints beyond the passive range of motion and do not result in joint cavitation."[32]p20 Note that these definitions do not specify the type of provider (e.g., physical therapist, chiropractor, primary care physician) performing the procedure; thus, the appropriateness ratings are applicable regardless of the practitioner. More detail on the RUAM is found in the RUAM manual.[3]

Analysis

The 1–9 appropriateness ratings given by each panelist after the face-to-face meeting were analyzed to generate one of three overall ratings for cervical mobilization and manipulation for each clinical scenario: appropriate, equivocal and inappropriate. The first analysis determined whether there was disagreement across the panelists' appropriateness ratings for any clinical scenario. Disagreement was defined as having at least four panelists' ratings in the 1–3 range and at least four in the 7–9 range. Agreement was defined as having at least 8 of the ratings in the 1–3, 4–6, or 7–9 range. If there was no disagreement and the median value of the ratings across the panel was 1–3, then the therapy was rated as inappropriate for that clinical scenario. If there was no disagreement and the median value of the ratings was 7–9, the therapy was rated as appropriate for that clinical scenario. The appropriateness for a therapy for a clinical scenario was rated as equivocal if: 1) 9 panelists gave a rating of 4, 5 or 6—i.e., there was agreement that benefits generally equaled risks; 2) panelists gave widely polarized ratings—i.e., there was disagreement; or 3) panelists' ratings were scattered across the scale—i.e., uncertainty as to appropriateness—and the median value was in the 4–6 range. The last two of these identify potential targets for future research.

The amount of disagreement, the dispersion of the ratings measured by the mean absolute deviation (MAD) from the median, and the proportions of clinical scenarios rated as appropriate, equivocal and inappropriate were compared between the at-home and in-meeting ratings. Comparisons of ratings between therapies and between scenarios with different histories of conservative care used paired t-tests, and comparisons of numbers of scenarios classified as appropriate versus inappropriate between therapies and between different histories of conservative care used χ^2 tests. Calculations of agreement and appropriateness were conducted using Microsoft Excel and Java.

We used decision tree analysis (DTA) to see if simplified rules could be identified regarding the elements (patient characteristics) that most affect the appropriateness of cervical mobilization and cervical manipulation.[33] DTA looks for the smallest number of patient characteristics or combinations of characteristics that can provide an accurate prediction of appropriate or inappropriate ratings. These simplified rules can provide information that is not always obvious from individual ratings across hundreds of clinical scenarios.

We first identified the set of patient characteristics (Supplemental Digital Content 1, Table A.1) that made up the clinical scenarios, and then defined each scenario as the presence or absence (or for some such as pain or imaging, a particular level) of each characteristic, and included all as predictor variables in the DTA. Since the clinical scenarios did not always mention all patient characteristics, we assumed that if a characteristic wasn't mentioned it was absent in that scenario. When predicting a rating of inappropriate we compared clinical scenarios with that rating to those without that rating—i.e., to those with ratings of either appropriate or equivocal. Similarly, in predicting a rating of appropriate, we compared those scenarios to those rated as inappropriate or equivocal. Some clinical scenarios were not included in the DTA because they were each made up of single patient characteristics that were considered as being added to an unspecified scenario "that would otherwise be rated as appropriate."[32] These clinical scenarios and their ratings are described separately (Supplemental Digital Content 1: Table A.2).

The DTA was conducted using the C4.5 algorithm[34] of the R statistical package (available at: https://cran.r-project.org/). In this algorithm tree branches are formed based on the characteristic that provides the most information gain at each step, and each branch is reached only by first meeting the characteristic of the branch before. The algorithm ends by returning to remove branches that are no longer useful.

The project was reviewed and determined to be exempt by RAND's Human Subjects Protection Committee.

Results

The panelists reported varying times for the ratings done at home, but 2–3 hours was roughly the norm. During discussions at the face-to-face meeting some revisions were made to the initial list of clinical scenarios to make them more clinically relevant and to make the groups of scenarios more homogeneous with respect to appropriateness. For example, "No history or signs of red flags" was considered preferable to "No clinical risk factors for

radiographic contraindications to cervical manipulation." The panelists then agreed on how to operationalize "red flags." Other changes included removing some unlikely scenarios, adding others to provide more detail, and changing some terminology—e.g., "Radiculopathy" was changed to "Neurologic findings" in multiple scenarios. Of the final set of 372 clinical scenarios for each therapy 330 were unchanged. More detail on the exact changes made is found in the published RAND report.[32]

Table 1 compares the initial at-home ratings to the final face-to-face ratings for the 330 unchanged scenarios and shows the ratings given, their dispersion (MAD) across panelists, and the number of clinical scenarios where there was agreement, a spread of ratings (uncertain), and clear disagreement. As can be seen, agreement increased, and dispersion and disagreement generally decreased between the two sets of ratings. Across the full final set of clinical scenarios, ratings were higher for mobilization than manipulation, and higher for either therapy when non-manual conservative care for the episode had failed than when it had not been tried (all paired t-tests p<.001).

Table 2 gives the number of clinical scenarios rated as appropriate, inappropriate or equivocal in the final face-to-face ratings across the final set of scenarios. More clinical scenarios were rated inappropriate than appropriate, and parallel to the results in Table 1, more scenarios were rated appropriate for mobilization than manipulation (χ^2 p<.05), and more, but not significantly more, were rated appropriate for both therapies when non-manual conservative care for the episode had failed. Between half and two-thirds of clinical scenarios were rated equivocal, and in most cases, this was due to a spread of ratings (uncertainty) with a median rating in the 4–6 range. Details on the ratings given to each clinical scenario are found in the published RAND report.[32]

DTA was applied to 278 of the 372 final clinical scenarios. Figures 1 and 2 show the results identifying the patient characteristics that best predict a clinical scenario being rated as inappropriate (versus appropriate or equivocal) for mobilization and manipulation, respectively. A table that presents the information shown in these figures in a format that might be easier to include in clinical guidelines is included as Supplemental Digital Content 1: Table A.3. The decision trees for each therapy both begin with the same three steps, but then become quite different. Both therapies are rated inappropriate in the presence of red flags (factors where the risk may outweigh the benefit, such as: fever greater than 100 degrees F; prolonged corticosteroid use; unexplained weight loss; history of cancer; history of serious systemic inflammatory arthritides or vasculitides; and endocrinopathies that affect calcium metabolism). Both therapies were also rated inappropriate for those without red flags if they had unfavorable prior experience with manual therapy and major neurologic findings (at least one of the following: neurologic signs of cervical myelopathy; progressive unilateral muscle weakness and/or motor loss documented by repeat exam over time; sensory deficits other than related to dermatomes or peripheral nerves; and/or electrodiagnostic findings of acute and/or progressive radiculopathy). Mobilization was found to be inappropriate for one other group: those without red flags or major neurologic findings who have had an unfavorable prior experience with manual therapy, have not tried any non-manipulative therapy for this episode, and had test findings of cervical disc herniation, stenosis, or foraminal osteophytosis.

On the other hand, manipulation was rated as inappropriate under five other conditions. Three for patients without red flags or major neurologic findings but with unfavorable prior experience with manual therapy: 1) those with clinically substantial traumatic etiology, and no signs of painful or limited range of motion; 2) those with clinically substantial traumatic etiology and painful or limited range of motion, but where no additional testing was done; and 3) those with no or minimal traumatic etiology who have not tried any non-manual therapy for this episode. The two other conditions where manipulation was rated inappropriate were for patients without red flags but with major neurologic findings where non-manual conservative care had not been tried: 1) those with no prior experience with manual therapy and clinically substantial traumatic etiology; and 2) those with no response to previous manual therapy and no or minimal traumatic etiology.

The decision tree analysis was fairly accurate. Only 9 clinical scenarios out of 278 for mobilization were misclassified as appropriate or equivocal when they were actually rated as inappropriate, and no clinical scenario was misclassified as inappropriate when it was actually rated as appropriate or equivocal (overall a 3.2% error rate). The same numbers for manipulation were 5 and 5 (3.6% error rate).

We also performed DTA predicting ratings of appropriate, versus inappropriate or equivocal, for both therapies. It was more complex to predict appropriateness than inappropriateness: the decision trees were less accurate and required inclusion of more patient characteristics. Seven clinical scenarios out of 278 for mobilization were misclassified as appropriate when they were actually rated as inappropriate or equivocal, and 4 clinical scenarios were misclassified as inappropriate or equivocal when they were actually rated as appropriate (overall a 4.0% error rate). The same numbers for manipulation were 6 and 8 (5.0% error rate).

Table 3 shows the percent of clinical scenarios each patient characteristic helped classify and the direction of its influence for both therapies and both predictions. As can be seen, the presence of red flags, as the first split in the tree, was the most useful (100%) to the accurate prediction of an inappropriate rating Unfavorable prior experience with cervical manual therapy was next most useful for both therapies because this characteristic provided information on inappropriateness for 86.3% (1 – 38/278) of scenarios—i.e., all those without red flags that made it past the first filter. Major neurologic findings were useful for identifying 18.7% (52/278) of inappropriate clinical scenarios for mobilization because only 52 scenarios (16 + 18 + 12 + 6 in Figure 1) made it past the unfavorable prior experience filter for this consideration. Whereas, major neurologic findings were useful for 86.3% of scenarios for manipulation because all those that made it past the red flags filter, no matter their prior experience, were subject to this consideration. The patient characteristics most useful to the accurate prediction of an appropriate rating were the absence of cervical disc herniation, stenosis, or foraminal osteophytosis on additional testing (100%). The number of patient characteristics involved and substantial size of the usefulness percentages indicate the complexity of each prediction.

As mentioned above, 94 clinical scenarios were excluded from the DTA because they were each rated as being added to an unspecified scenario "that would otherwise be rated as

appropriate." Mobilization was rated inappropriate for 38 of these scenarios and appropriate for 5. The same numbers for manipulation were 55 and 1. All other scenarios were rated equivocal. Both therapies were rated inappropriate in the face of several serious patient signs —e.g., any brainstem neurologic findings; nystagmus or dizziness during or immediately after provocative testing; or clotting disorders with no or abnormal clotting or bleeding tests. There were also several clinical scenarios where manipulation was rated inappropriate and mobilization was rated equivocal. Examples of these include: clinical or physical examination evidence of occlusive vascular disease, hypertension, or history of transient ischemic attack of carotid origin. The full list of these clinical scenarios and their ratings are found in Supplemental Digital Content 1: Table A.2.

Discussion

This study used a well-validated expert panel-based approach (the RUAM) to obtain ratings of the appropriateness of cervical mobilization and manipulation for various types of patients with CNP. The range of potential patients with CNP were represented by 372 clinical scenarios developed to cover all possible presentations. Of these 65 were rated as appropriate, 115 were rated as inappropriate, and 192 were rated as equivocal (agreement that benefits roughly equal risks, or uncertainty as to whether benefits are greater or less than risks with a median rating of rough equality) for cervical mobilization. The numbers for cervical manipulation were 42 appropriate, 146 inappropriate, and 184 equivocal. Analysis of these ratings using decision tree analysis indicated that the presence of red flags was the main determinant of a rating of inappropriate for both therapies. About half of the clinical scenarios received a rating of equivocal because of a lack of agreement in ratings across panelists with a median rating in the 4–6 range (benefits roughly equal to risks).

As would be expected from the controversies around the safety and use of cervical manipulation,[19-23] more clinical scenarios were rated inappropriate than appropriate for both therapies, and more were rated inappropriate for cervical manipulation than mobilization. A study by Puentedura et al[27] rated almost half (44.8%) of 134 cases of severe adverse events associated with cervical manipulation as preventable if clinicians had heeded contraindications and red flags. In our study, the contraindications and red flags used in the Puentedura et al study, including unfavorable response to previous manual therapy, were all associated with cervical manipulation being rated as inappropriate.

Note that the numbers of clinical scenarios rated appropriate, equivocal, and inappropriate bear no relationship to the number of patients receiving each type of care.

For example, although half the clinical scenarios were rated equivocal, the proportion of patients who present with these clinical scenarios may be lower or higher. A future article from this study will present the results of these ratings applied to the healthcare records of a representative sample of patients using chiropractic care for CNP to determine the proportion of patients that present with each clinical scenario and the proportion who are receiving appropriate and inappropriate care. Nevertheless, information on the clinical scenario rated inappropriate is still useful for guideline development.

Considerable attention has been given to the relationships between psychosocial stress and the etiology and treatment of back pain [35] and these variables help define stratified and stepped care plans.[36, 37] In fact, decision tree analysis of the results of a similar expert panel for chronic low back pain found that biomechanical or psychosocial stress was a useful characteristic to predict the appropriateness of spinal mobilization and manipulation for a clinical scenario.[38] However, it appears that psychosocial stress (i.e., depression requiring drug treatment, alcohol or narcotic dependence, recent suicide attempt, severe anxiety, evidence of stressful life situation such as bereavement, job change, job or family dissatisfaction, or litigation or compensation issues) was not useful in predicting either the appropriateness or inappropriateness of mobilization or manipulation for CNP scenarios.

This study used an internationally-recognized and well-validated method to translate available evidence and expert clinical acumen into appropriateness ratings across an exhaustive list of 372 clinical scenarios which could present as CNP. However, the approach is not without limitations. Panelists were presented with a full synthesis of all available evidence on the safety and effectiveness/efficacy of cervical mobilization and manipulation for CNP. However, the available evidence has gaps, including the facts that clinical trials only include a subset of patients with CNP and the analyses of trial data don't present results by distinct clinical scenarios. Therefore, panelists' clinical acumen was essential to the process and not without its own biases. For example, even though the evidence available and presented on cervical manipulation showed greater efficacy with no more adverse events than mobilization,[18] panelists still rated mobilization higher (as more appropriate and having a better benefit to risk ratio) than manipulation.

Panelists were also asked to rate the appropriateness of 372 clinical scenarios for each therapy, and it is difficult to perform so many ratings consistently and without error. We know of one small set of clinical scenarios where the panel produced what seem to be inconsistent results for manipulation, rating it as inappropriate (median rating of 3) for 4 scenarios where there was non-traumatic or minimally traumatic etiology and as equivocal (median rating of 4) for 4 scenarios where there was a clinically substantial traumatic etiology. According to the pattern of ratings seen in related, near-by scenarios, these 8 scenarios should have received the same rating—either all inappropriate or all equivocal. These clinical scenarios were presented to panelists as part of two sets: 36 scenarios with non-traumatic or minimal traumatic etiology and 36 with clinically substantial traumatic etiology. Within each set panelists worked through different levels of prior experience with manual therapy from no experience to favorable, no, and unfavorable responses. Although in general slightly lower ratings were given for scenarios with a clinically substantial traumatic etiology, consistent with what was shown in Table 3, prior experience with manual therapy had a larger impact on ratings. It seems that when moving from scenarios with a favorable response to prior manual therapy to those with no response panelists made bigger ratings reductions for the first set of scenarios (non-traumatic or minimally traumatic) than for the second (clinically substantial trauma) and these resulted in the slight but critical difference in median ratings. Nevertheless, the accuracy of the DTA's predictions was one indication of at least the internal consistency of the ratings.

A well-validated expert panel-based approach was used to develop and then rate the appropriateness of the use of cervical mobilization and manipulation across an exhaustive list of clinical scenarios which could present for CNP. In keeping with the controversy surrounding the safety of cervical mobilization and manipulation, more clinical scenarios were rated inappropriate than appropriate for both, but the number rated as inappropriate for cervical manipulation was higher. Half of the scenarios were rated equivocal due to panelists giving a range of ratings whose median was in the range of rough equivalence. If these scenarios turn out to represent a substantial number of patients seen, this last could be a worthwhile target for future research. Nonetheless, all clinical scenarios that included red flags, and some others involving major neurologic findings, especially if there was an unfavorable response to prior manual therapy, were found to be rated as inappropriate for both cervical mobilization and manipulation. This information should be used to inform clinical guidelines recommending these therapies for patients with chronic neck pain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Brook R Appropriateness: The next frontier. Br Med J 1994;308:218. [PubMed: 8111255]

2. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J and Park RE. A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care 1986;2:53–63. [PubMed: 10300718]

3. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader JP and Kahan JP RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND Corporation, 2001.

4. McClellan M and Brook RH. Appropriateness of care: A comparison of global and outcome methods to set standards. Med Care 1992:565–586. [PubMed: 1614228]

5. Shekelle P The appropriateness method. Med Decis Making 2004;Mar-Apr:228–231. [PubMed: 15090107]

6. Merrick NJ, Fink A, Park RE, Brook RH, Kosecoff J, Chassin MR and Solomon DH. Derivation of clinical indications for carotid endarterectomy by an expert panel. Am J Public Health 1987;77:187–190. [PubMed: 3799858]

7. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ and Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. New Engl J Med 1998;338:1888–1895. [PubMed: 9637810]

8. Tobacman JK, Scott IU, Cyphert S and Zimmerman B. Reproducibility of measures of overuse of cataract surgery by three physician panels. Med Care 1999;37:937–945. [PubMed: 10493471]

9. Selby JV, Fireman BH, Lundstrom RJ, Swain BE, Truman AF, Wong CC, Froelicher ES, Barron HV and Hlatky MA. Variation among hospitals in coronary-angiography practices and outcomes after myocardial infarction in a large health maintenance organization. New Engl J Med 1996;335:1888–1896. [PubMed: 8948565]

10. Normand S-LT, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD and McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54:387–398. [PubMed: 11297888]

11. Shekelle PG, Chassin MR and Park RE. Assessing the predictive validity of the RAND/UCLA appropriateness method criteria for performing carotid endarterectomy. Int J Technol Assess Health Care 1998;14:707–727. [PubMed: 9885461]

12. Hemingway H, Crook AM, Feder G, Banerjee S, Dawson JR, Magee P, Philpott S, Sanders J, Wood A and Timmis AD. Underuse of coronary revascularization procedures in patients considered appropriate candidates for revascularization. New Engl J Med 2001;344:645–654. [PubMed: 11228280]

13. Kravitz RL, Laouri M, Kahan JP, Guzy P, Sherman T, Hilborne L and Brook RH. Validity of criteria used for detecting underuse of coronary revascularization. JAMA 1995;274:632–638. [PubMed: 7637144]

14. Johannes CB, Le TK, Zhou X, Johnston JA and Dworkin RH. The prevalence of chronic pain in United States adults: results of an Internet-based survey. J Pain 2010;11:1230–1239. [PubMed: 20797916]

15. Webb R, Brammah T, Lunt M, Urwin M, Allison T and Symmons D. Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. Spine 2003;28:1195–1202. [PubMed: 12782992]

16. Hurwitz EL, Carragee EJ, van der Velde G, Carroll LJ, Nordin M, Guzman J, Peloso PM, Holm LW, Côté P and Hogg-Johnson S. Treatment of neck pain: noninvasive interventions: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. J Manip Physiol Ther 2009;32:S141–75.

17. Hurwitz EL. Epidemiology: spinal manipulation utilization. J Electromyography Kinesiol 2012;22:648–54.

18. Coulter ID, Crawford C, Hurwitz EL, Vernon H, Khorsan R, Booth MS and Herman PM. Manipulation and mobilization for treating chronic neck pain: a systematic review and meta-analysis. Pain Physician 2019;[in press].

19. Smith W, Johnston S, Skalabrin E, Weaver M, Azari P, Albers G and Gress D. Spinal manipulative therapy is an independent risk factor for vertebral artery dissection. Neurology 2003;60:1424–1428. [PubMed: 12743225]

20. Wand BM, Heine PJ and O'Connell NE. Should we abandon cervical spine manipulation for mechanical neck pain? Yes. BMJ 2012;344:e3679. [PubMed: 22677796]

21. Cassidy JD, Bronfort G and Hartvigsen J. Should we abandon cervical spine manipulation for mechanical neck pain? No. BMJ: British Medical Journal (Online) 2012;344.

22. Ernst E Adverse effects of spinal manipulation: a systematic review. J R Soc Med 2007;100:330–338. [PubMed: 17606755]

23. Nielsen SM, Tarp S, Christensen R, Bliddal H, Klokker L and Henriksen M. The risk associated with spinal manipulation: an overview of reviews. Systematic reviews 2017;6:64. [PubMed: 28340595]

24. Cassidy JD, Boyle E, Côté P, He Y, Hogg-Johnson S, Silver FL and Bondy SJ. Risk of vertebrobasilar stroke and chiropractic care: results of a population-based case-control and case-crossover study. J Manipulative Physiol Ther 2009;32:S201–S208. [PubMed: 19251066]

25. Cassidy JD, Boyle E, Côté P, Hogg-Johnson S, Bondy SJ and Haldeman S. Risk of carotid stroke after chiropractic care: a population-based case-crossover study. J Stroke Cerebrovasc Dis 2017;26:842–850. [PubMed: 27884458]

26. Haldeman S, Kohlbeck FJ and McGregor M. Stroke, cerebral artery dissection, and cervical spine manipulation therapy. J Neurol 2002;249:1098–1104. [PubMed: 12195461]

27. Puentedura EJ, March J, Anders J, Perez A, Landers MR, Wallmann HW and Cleland JA. Safety of cervical spine manipulation: are adverse events preventable and are manipulations being performed appropriately? A review of 134 case reports. J Man Manip Ther 2012;20:66–74. [PubMed: 23633885]

28. Brook RH. Assessing the appropriateness of care—its time has come. JAMA 2009;302:997–998. [PubMed: 19724048]

29. Nair R, Aggarwal R and Khanna D Methods of formal consensus in classification/diagnostic criteria and guideline development Semin Arthritis Rheum. Elsevier, 2011:95–105.

30. Coulter ID, Hurwitz EL, Adams AH, Meeker WC, Hanson D, Mootz W, Aker P, Genovese BJ and Shekelle PG The Appropriateness of Spinal Manipulation and Mobilization of the Cervical Spine: Literature Review, Indications and Ratings by a Multidisciplinary Expert Panel. Santa Monica, CA: RAND Corporation, 1995.

31. Coulter ID, Shekelle PG, Mootz RD and Hansen DT. The use of expert panel results: The RAND panel for appropriateness of manipulation and mobilization of the cervical spine. Journal of Topics in Clinical Chiropractic 1995;2:54–62.

32. Coulter ID, Whitley MD, Vernon H, Hurwitz EL, Shekelle PG and Herman PM Determining the Appropriateness of Spinal Manipulation and Mobilization for Chronic Neck Pain Indications and Ratings by a Multidisciplinary Expert Panel. Santa Monica, CA: RAND Corporation, 2018.

33. Hastie T, Tibshirani R and Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2016.

34. Quinlan JR. C4.5: Programs for Machine Learning. Burlington, MA: Morgan Kaufmann Publishers, 1993.

35. Turk DC, Fillingim RB, Ohrbach R and Patel KV. Assessment of psychosocial and functional impact of chronic pain. J Pain 2016;17:T21–T49. [PubMed: 27586830]

36. Von Korff M and Moore JC. Stepped care for back pain: activating approaches for primary care. Ann Intern Med 2001;134:911–917. [PubMed: 11346328]

37. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, Konstantinou K, Main CJ, Mason E and Somerville S. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. The Lancet 2011;378:1560–1571.

38. Herman PM, Hurwitz EL, Shekelle PG, Whitley MD and Coulter ID. Clinical Scenarios for which Spinal Mobilization and Manipulation Are Considered by an Expert Panel to be Inappropriate (and Appropriate) for Patients with Chronic Low Back Pain. Med Care 2019;57:391–398. [PubMed: 30870390]
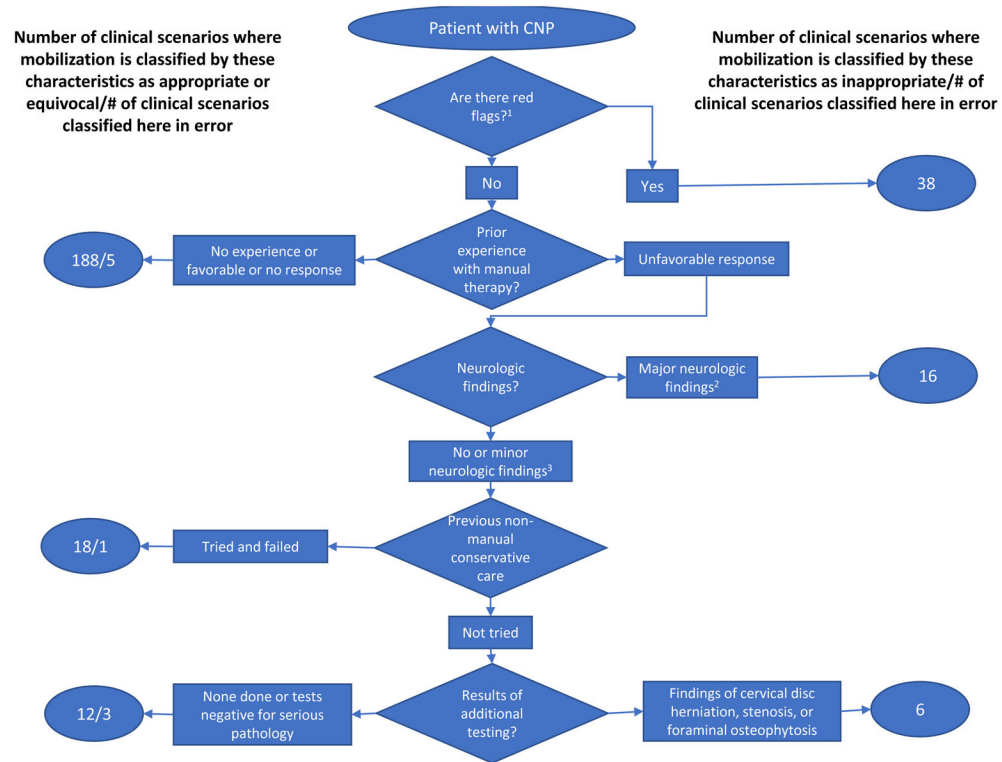
Figure 1.

Results of the decision tree analysis of the patient characteristics that best define the clinical scenarios where mobilization was rated inappropriate

CNP = chronic neck (cervical) pain

[1]Red flags = Those factors where the risk may outweigh the benefit, such as: fever greater than 100 degrees F; prolonged corticosteroid use; unexplained weight loss; history of cancer; history of serious systemic inflammatory arthritides or vasculitides; endocrinopathies that affect calcium metabolism.

[2]Major neurologic findings = A At least one of the following: neurologic signs of cervical myelopathy; progressive unilateral muscle weakness and/or motor loss documented by repeat exam over time; sensory deficits other than related to dermatomes or peripheral nerves; and/or electrodiagnostic findings of acute and/or progressive radiculopathy.

[3]Minor neurologic findings = At least one of the following: asymmetrically decreased reflexes in upper extremity; documented dermatomal or peripheral nerve sensory changes which may include deficit, paresthesia, and hyperesthesia; non-progressive unilateral muscle weakness and/or parasthesia that follows a radicular pattern.

Figure 2.
Results of the decision tree analysis of the patient characteristics that best define the clinical scenarios where manipulation was rated inappropriate

CNP = chronic neck (cervical) pain

[1]Red flags = Those factors where the risk may outweigh the benefit, such as: fever greater than 100 degrees F; prolonged corticosteroid use; unexplained weight loss; history of cancer; history of serious systemic inflammatory arthritides or vasculitides; endocrinopathies that affect calcium metabolism.

[2]Major neurologic findings = A At least one of the following: neurologic signs of cervical myelopathy; progressive unilateral muscle weakness and/or motor loss documented by repeat exam over time; sensory deficits other than related to dermatomes or peripheral nerves; and/or electrodiagnostic findings of acute and/or progressive radiculopathy.

[3]Minor neurologic findings = At least one of the following: asymmetrically decreased reflexes in upper extremity; documented dermatomal or peripheral nerve sensory changes which may include deficit, paresthesia, and hyperesthesia; non-progressive unilateral muscle weakness and/or parasthesia that follows a radicular pattern.
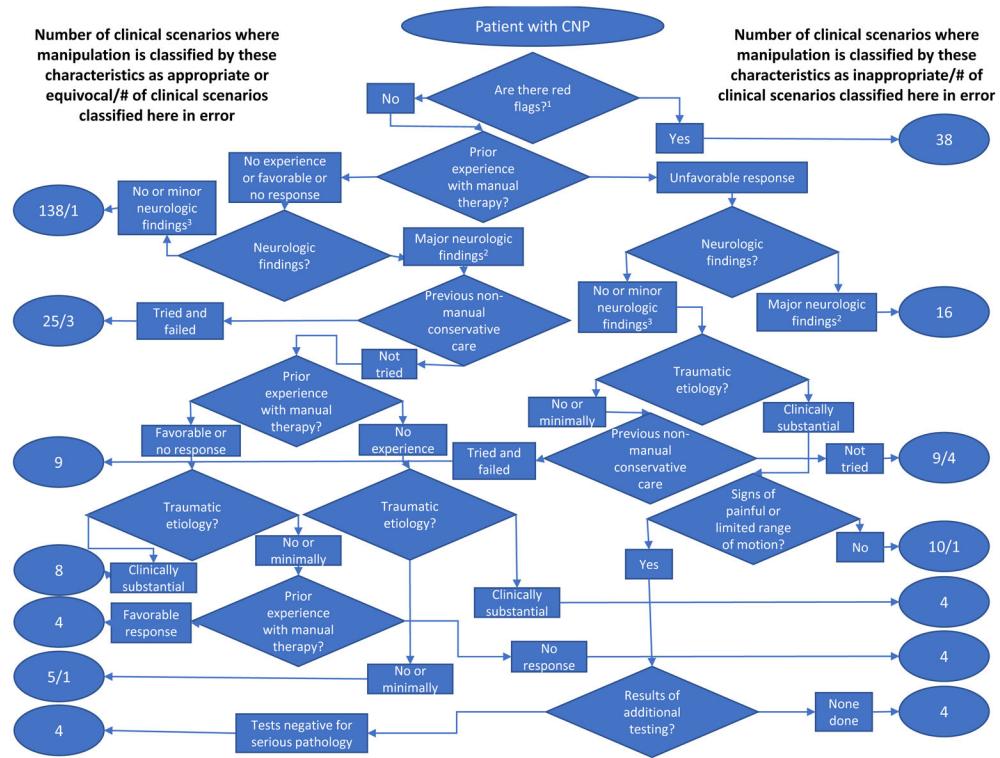
Table 1.

Change in median ratings and extent of agreement and disagreement between the initial at-home ratings and final face-to-face meeting ratings by panelists using the RAND/UCLA Appropriateness Method

Item	Cervical Mobilization				Cervical Manipulation			
	No other adequate conservative care		Nonmanipulative conservative care has failed		No other adequate conservative care		Nonmanipulative conservative care has failed	
	Initial Ratings	Final Ratings	Initial Ratings	Final Ratings	Initial Ratings	Final Ratings	Initial Ratings	Final Ratings
For the clinical scenarios that did not change between rounds (330 for each cervical mobilization and manipulation)								
Average Median	4.7	4.4	5.2	4.4	3.9	3.9	4.2	4.2
MAD from Median	1.5	1.2	1.4	1.2	1.5	1.2	1.6	1.2
Number (%) Agreement	22 (13.3%)	61 (37.0%)	44 (26.7%)	72 (43.6%)	38 (23.0%)	55 (33.3%)	44 (26.7%)	64 (38.8%)
Number (%) Uncertain	138 (83.6%)	102 (61.8%)	117 (70.9%)	89 (53.9%)	122 (73.9%)	109 (66.1%)	115 (69.7%)	100 (60.6%)
Number (%) Disagreement	5 (3.0%)	2 (1.2%)	4 (2.4%)	4 (2.4%)	5 (3.0%)	1 (0.6%)	6 (3.6%)	1 (0.6%)
For all clinical scenarios used in the final round (372 for each cervical mobilization and manipulation)								
Average Median		4.3		4.7		3.9		4.2
MAD from Median		1.2		1.2		1.3		1.2
Number (%) Agreement		70 (37.6%)		82 (44.1%)		63 (33.9%)		75 (40.3%)
Number (%) Uncertain		114 (61.3%)		100 (53.8%)		122 (65.6%)		110 (59.1%)
Number (%) Disagreement		2 (1.1%)		4 (2.2%)		1 (0.5%)		1 (0.5%)

MAD = mean absolute deviation

Table 2.

Final appropriateness ratings across the final list of clinical scenarios by panelists using the RAND/UCLA Appropriateness Method

Rating	Cervical Mobilization				Cervical Manipulation			
	No other adequate conservative care		Nonmanipulative conservative care has failed		No other adequate conservative care		Nonmanipulative conservative care has failed	
	Number	%	Number	%	Number	%	Number	%
Inappropriate	64	34.4%	51	10.0%	80	43.0%	66	35.5%
Equivocal	95	51.1%	97	65.8%	90	48.4%	94	50.5%
Agreement and equivocal	10	5.4%	19	10.2%	4	2.2%	7	3.8%
Disagreement	2	1.1%	4	2.2%	1	0.5%	1	0.5%
Uncertain and equivocal	83	44.6%	74	39.8%	85	45.7%	86	46.2%
Appropriate	27	14.5%	38	20.4%	16	8.6%	26	14.0%
Total	186	100%	186	100%	186	100%	186	100%

Table 3.

The usefulness of each patient characteristic in predicting whether a clinical scenario containing that characteristic was rated inappropriate or appropriate for each therapy (i.e., the percent of scenarios each characteristic helps classify)

Patient characteristics*	Predicting a rating of inappropriate**				Predicting a rating of appropriate**			
	Mobilization		Manipulation		Mobilization		Manipulation	
Presence of red flags	100.0%	↗	100.0%	↗	65.3%	↘	35.0%	↘
Prior unfavorable experience with spinal manual therapy	86.3%	↗	86.3%	↗	51.6%	↘	61.0%	↘
Major neurologic findings	18.7%	↗	86.3%	↗	41.5%	↘	27.1%	↘
Previous non-manual conservative care failed	13.0%	↘	24.5%	↘	---		---	
Additional (e.g., advanced imaging) testing show cervical disc herniation, stenosis, or foraminal osteophytosis	6.5%	↗	2.9% ***	↘	100.0%	↘	100.0%	↘
Clinically substantial traumatic etiology	---		21.9%	↗	30.7%	↘	23.5%	↘
No signs of painful/limited active range of motion	---		4.7%	↗	---		---	
Radiographs showing advanced spinal degeneration	---		---		26.0%	↘	65.3%	↘
No cervical nerve root radiculopathy	---		---		7.6%	↘	---	
Joint dysfunction in upper cervical spine	---		---		---		14.8%	↘
No local pathology	---		---		---		11.9%	↗
Continued psychosocial stress	---		---		---		---	

*
The patient characteristics are all fully defined in the Appendix.

**
Predictions of a rating of inappropriate were versus ratings of equivocal or appropriate. Predictions of appropriate were versus ratings of equivocal or inappropriate.

***
Tests negative for serious pathology included here with other test results.

↗ = This patient characteristic predicts this rating.

↘ = This patient characteristic predicts against this rating.