Analysis of scientific truth status in controlled rehabilitation trials

Roger Kerry FMACP, MCSP, MSc,¹ Aurélien Madouasse DVM, PhD, ²* Antony Arthur PhD³ and Stephen D. Mumford PhD⁴

¹Associate Professor, Division of Physiotherapy Education, University of Nottingham, Nottingham, UK
²Research Fellow, School of Veterinary Medicine, University of Nottingham, Nottingham, UK
³Associate Professor, Division of Nursing, University of Nottingham, Nottingham, UK
⁴Professor, Department of Philosophy, University of Nottingham, Nottingham, UK

Keywords
evidence-based practice, philosophy, randomized controlled trials, rehabilitation, systematic reviews, truth

Abstract

Rationale, aims and objectives Systematic reviews, meta-analyses and clinical guidelines (reviews) are intended to inform clinical practice, and in this sense can be thought of as scientific truthmakers. High-quality controlled trials should align to this truth, and method quality markers should predict truth status. We sought to determine in what way controlled trial quality relates to scientific truth, and to determine predictive utility of trial quality and bibliographic markers.

Method A sample of reviews in rehabilitation medicine was examined. Two scientific truth dimensions were established based on review outcomes. Quality and bibliographic markers were extracted from associated trials for use in a regression analysis of their predictive utility for trial truth status. Probability analysis was undertaken to examine judgments of future trial truth status.

Results Of the 93 trials included in contemporaneous reviews, overall, \( n = 45 \) (48%) were true. Randomization was found more in true trials than false trials in one truth dimension \( (P = 0.03) \). Intention-to-treat analysis was close to significant in one truth dimension \( (P = 0.058) \), being more commonly used in false trials. There were no other significant differences in quality or bibliographic variables between true and false trials. Regression analysis revealed no significant predictors of trial truth status. Probability analysis reported that the reasonable chance of future trials being true was between 2 and 5%, based on a uniform prior.

Conclusions The findings are at odds with what is considered gold-standard research methods, but in line with previous reports. Further work should focus on scientific dynamics within healthcare research and evidence-based practice constructs.

Introduction

Contradictory research findings from randomized controlled trials have previously been reported [1]. Further, models demonstrating that most claimed research findings are false have been given [2]. Such models claim that falsity may simply be a measure of prevailing study bias. Judgment of trial quality is made based on indicators of bias control: journals demand adherence of trials to accepted quality criteria (e.g. CONSORT [3]); reviewers score trials based on accepted quality criteria (e.g. Cochrane [4]); and consumer cognizance of trial quality is encouraged [5]. Ioannidis reported that the effect of contradicted findings was seen most clearly in studies without randomization or homogeneous grouping [1]. Despite this apparent importance on trial quality variables, there is some concern over whether or not high-quality trials (defined by its considered ability to control for bias) do actually produce the most truthful findings [6–9]. More broadly, there has been increasing concern about the nature and trajectory of science within health research and evidence-based practice, for example, see Fallis, Mittra, Tonelli, and Cartwright and Munro [10–13].

This paper adapts Ioannidis’ approach of comparing trials to each other singularly, to compare individual trial findings with those of their parent systematic reviews. A function of systematic reviewing is to tolerate variance of trial findings and account for
discrepancies by considering the methodological qualities of individual trials. Thus, a systematic review should produce a form of quality-weighted outcome consistent with the outcomes of the best-quality trials. A clinically meaningful truth can be established (systematic review outcome) to which individual trials may be compared. We refer to this truth as a scientific truth. Although this definition may be contentious, it is presented as nothing more than an indication of knowledge status at a single time point which represents the current state of scientific knowledge – that is, that knowledge which is a result of scientific investigation. This in turn aligns well with the principles of evidence-based practice which attributes a strong level of evidence to high-quality reviewing of primary data, the overall intent being that such findings should influence clinical decision making.

This study was conducted to identify truth status of published trials, and report differences in quality variables between true and false trials. We hypothesized that better quality trials are more likely to be true, and quality variables should predict truth status of trials. Further, we report on probabilistic judgments of the likelihood of future trials being true. The analysis is of existing data which have been established and utilized through existing research methods. No claims or comments are made on these methods. The aims of the study were to merely map out decisions and commitments which have already been established within the science, for example, trial quality has already been scored and implemented within a review. The present study is simply tracking relationships of trial and review outcomes within the existing science. The study is based on an existing protocol [1].

Methods

Eligible studies

A sample of recent (2007–2010) systematic reviews associated with specific areas of rehabilitation medicine was identified through their inclusion in current clinical guidelines and by searching: The Database of Abstracts and Reviews of Effects; Physiotherapy Evidence Database; PubMed; Web of Science; Cochrane Library. The five most recent systematic reviews with the largest number of controlled trials were included in this study. The trials from each of these reviews were used as the base units for this analysis.

Classification of reviews and trials

In accordance with existing protocol [1], the outcome from each systematic review was recorded based on the final interpretation by the authors in the ‘Abstract’ and ‘Discussion’ sections of the reviews. Outcomes were classified as ‘negative’ if the review claimed that the experimental intervention was no better than the control or comparators, and ‘positive’ if the experimental intervention was better than the control or comparators. Outcomes from each trial were classified in the same way. Truth was then established in two ways: first, ultimate truth (ULT) was judged to be the outcome of the parent systematic review. Real-time truth (RTT) was established by re-systematically reviewing trials at the time point of publication of each trial. The re-review process was identical to the process reported in the final associated published systematic review. Thus, a different systematic review outcome was produced at the point of publication of each individual trial. This was the RTT. The purpose of establishing two truth dimensions was to allow comparison of individual trials with as much clinical meaningfulness as possible, and to be inclusive of contradictory findings: ULT reflects what we know to be true at this moment in time; RTT reflects the dynamic status of scientific knowledge which changes in response to scientific evidence over time.

Each trial was then classified as being true or false in two ways: ULT status was established by the agreement between individual trial outcome and its parent systematic review outcome. If the outcomes agreed, the trial was classed as ULT-True, and if the outcomes differed, the trial was classed as ULT-False. RTT status was established by similar criteria, but considering the agreement between individual trial outcome and the re-review outcome at the point immediately prior to the publication of that particular trial.

Extraction of variables and temporal trends

The dependent variables were the truth status of each trial (ULT and RTT). In addition to the year of publication, independent variables were extracted based on the methodological quality criteria used by the parent systematic review. The exact criteria differed between the reviews. Common quality variables were randomization, use of intention-to-treat (ITT) analysis, allocation concealment, sample size, and blinding. Additionally, an overall quality of method measure was extracted from each review which was the score given to the trial by the authors of its parent systematic review. Bibliometric variables were recorded for each trial based on data available from Institute of Scientific Information databases. These were number of citations for each trial and the impact factor of journal. All variables were categorized into binary outcomes to reflect clinical interpretation of evidence, that is, either this is good evidence or it is not. Binary division points were established relevant to the variable.

Statistical analysis

Outcome prediction analysis was performed using PASW Statistics 18.0 (IBM SPSS Inc, Chicago, IL, USA). Statistical significance was considered for P-values <0.05 (two tailed). Temporal trends were described for truth status and overall quality of trials. Variable extraction took place and comparisons between true and false groups were made for each variable using Fisher exact test for binary variables and Mann–Whitney U-test for continuous variables. Outcome prediction using binary logistic regression examined associations between dependent variables (outcomes) of ULT and RTT, the binary state independent variables (exposures) of interest based on a priori assumptions (randomization, ITT analysis and blinding), and other independent variables (other exposures). A question-led, forced-entry, model-building strategy was used to establish which model best predicts outcome and best explains outcome variation. Beta values and standard errors, regression coefficients, P-values, and odds ratios for exposure (Expb) with 95% confidence intervals were reported for each variable. The overall fit of the model was considered by reporting -2xlog-likelihood (~2LL) and its associated chi-square statistic ($\chi^2$). Coefficient of determination ($R^2$) was reported as a measure of exposure influence on outcome variability.
Further prediction analysis was performed using WinBUGS 1.4.3 (Imperial College, MRC, UK [14]). A framework was used with a range of prior beta distributions for model parameters representing assumed clinical opinions (calculated via a custom-written function in R [15]). Parameters $\alpha$ and $\beta$ for beta distributions that reflect prior probabilities were calculated as a function of the desired mean and variance using the following formulas:

$$
\alpha = (\mu^2 - \mu^3 - \sigma \mu^2) / \sigma^2 \\
\beta = (2 \mu - \mu^2 + \mu^3 - \mu \sigma^2 + \mu \sigma^3) / \sigma^2 
$$

where $\mu$ and $\sigma$ were assumed mean and variance reflecting a theoretical range of clinical beliefs in success of trials: sceptic $\mu = 0.15$, $\sigma = 0.1$; uniform $\mu = 1$, $\sigma = 1$; enthusiastic $\mu = 0.85$, $\sigma = 0.1$.

Sampling distribution (number of successes of trials being true and total number of trials) was informed by the outcomes of the initial frequency analysis. Modelling was designed to estimate posterior estimates (predictive distribution) and the probability of exceeding a critical threshold – that is, the probability of at least 60% of future trials being true. Sixty per cent was chosen to reflect a conservative majority. Model parameters $p$ and $\eta$ for binomial posterior distributions were estimated using Markov chain Monte Carlo (MCMC) considered by the following model:

$$
Model: \quad y \sim \text{dbin}(p, n) \\
Y_{\text{new}} \sim \text{dbin}(p, 93) \\
S < - \text{step}(Y_{\text{new}} - 56) \\
Prior: \quad p \sim \text{dbeta}(a, b) \\
a < - \alpha \\
b < - \beta \\
Data: \quad (n, y)
$$

where $\sim$: stochastic (probabilistic) relationship; $<-$: deterministic (logical) relationship; $y$: outcome, that is, number of successes; $Y_{\text{new}}$: critical threshold, that is, number of success in a future 93 trials; $S$: the number of times at least 60% of future trials were successful; $p$: prior beta distribution; $a$ and $b$: beta parameters derived from (1); $n$: number of observed trials from data; and $y$: number of observed successes from data. Model parameters were estimated using MCMC as implemented in WinBUGS. Three chains using randomly generated starting values were run. Starting values were not found to influence posterior estimates. Model convergence was examined using informal visual assessment of the chains and a Gelman–Rubin convergence diagnostic [16]. Analyses used a burn-in of at least 250 iterations. All models converged well ahead of the end of this burn-in. Analysis was then based on an additional 1000 iterations.

**Results**

**Eligible studies**

Four hundred fourteen published rehabilitation systematic reviews (with or without meta-analysis) were identified between 2007 and 2010. Less than 100 of these met criteria of either being included in contemporaneous clinical guidelines and being based predominantly on controlled trials. Of these, less than half were based on more than 10 trials. The five reports with the highest number of associated trials were identified for use as a sample for this study. Clinical areas, published year range and number of associated trials (n) were:

- Manipulation for neck pain [17], 1977–2009, $n = 22$ [18–39];
- Acupuncture for low back pain [40], 1980–2007, $n = 22$ [26,34,41–60];
- Stroke rehabilitation (differences between approaches) [61], 1970–2005, $n = 13$ [62–74];
- Physiotherapy interventions for carpal tunnel syndrome [75], 2002–2009, $n = 26$ [76–101];
- Manipulation for low back pain [102], 1975–2007, $n = 10$ [103–112].

These 93 trials formed the base units of analysis.

**Classification of reviews and trials**

The outcomes of the five systematic reviews/guidelines were classified as either positive or negative and this classification was repeated by a blinded second reviewer. All reviews were positive (i.e. supported the intervention under trial), all reporting similar small/moderate effect sizes. There was 100% inter-rater agreement in these judgments. The outcome of each associated trial was classified and repeated with blinded second reviewer classification, again with 100% agreement. ULT and RTT were calculated with ‘true’ proportions being: manipulation for neck pain ULT 50%, RTT 32%; acupuncture for low back pain ULT 68%, RTT 55%; stroke rehabilitation ULT 54%, RTT 62%; carpal tunnel syndrome ULT 35%, RTT 46%; manipulation for low back pain ULT 40%, RTT 50%. The average measures were ULT: 49%; RTT: 47%. That is, just under half of all trials were true, in both truth dimensions (Table 1).

**Extraction of variables, temporal trends and prediction analysis**

Once ULT and RTT status had been established, pooling of data was permissible as the only relevant characteristic of each trial was its truth status. Independent variables concerning trial quality and bibliometrics were recorded and grouped to show proportions between true and false groups (Table 2).

With regard to ULT, 46 trials were true and 47 trials were false. The median sample size was larger in the false trials ($n = 131$ compared to $n = 117$). Randomization was more common in the true trials, but only by 1. All other quality and bibliometric variables were higher in the false trials. None of the differences were statistically significant. ITT analysis difference was close to significant ($P = 0.058$), with a higher proportion in the false trials. For RTT, 44 trials were true and 49 trials were false. Median sample size was greater in the true trials ($n = 130$ compared to $n = 119$). All quality variables other than ITT were higher in the true trials. All bibliometric variables other than recent publication were higher in the true trials. The difference in randomization was the only statistically significant outcome ($P = 0.03$) in favour of true trials.
### Table 1: Outcome classification of all reviews, re-reviews and truth classification of all trials

<table>
<thead>
<tr>
<th>Clinical area</th>
<th>Systematic review and outcome</th>
<th>Trials</th>
<th>Outcome n (%): ULT (T/n %): Re-review outcome n (%): RTT T/n %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical manipulation for neck pain</td>
<td>Gross et al. [17]+</td>
<td>n = 22</td>
<td>11/22 (50) 11/22 (50) 10/22 (45) 7/22 (32)</td>
</tr>
<tr>
<td>Acupuncture for low back pain</td>
<td>Rubinstein et al. [40]+</td>
<td>n = 22</td>
<td>13/22 (59) 15/22 (68) 15/22 (68) 12/22 (55)</td>
</tr>
<tr>
<td>Stroke rehabilitation (differences between</td>
<td>Pollock et al. [61]+</td>
<td>n = 13</td>
<td>6/13 (46) 7/13 (54) 5/13 (38) 8/13 (62)</td>
</tr>
<tr>
<td>approaches)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiotherapy for carpal tunnel syndrome</td>
<td>Huisстede et al. [75]+</td>
<td>n = 26</td>
<td>10/26 (39) 9/26 (35) 8/26 (30) 12/26 (46)</td>
</tr>
<tr>
<td>Manipulation for low back pain</td>
<td>Savigny et al. [102]+</td>
<td>n = 10</td>
<td>4/10 (40) 4/10 (40) 6/10 (60) 5/10 (50)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>44/93 (47) 46/93 (49) 44/93 (47) 44/93 (47)</td>
</tr>
</tbody>
</table>

ULT, ultimate truth; RTT, real-time truth; +, outcome supports efficacy of intervention under trial; T, trial reports the same outcome as associated review.

### Table 2: Proportions of True and False trials for each quality variable under each truth dimension

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>True n = 46</th>
<th>False n = 47</th>
<th>P-value</th>
<th>True n = 44</th>
<th>False n = 49</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size, median (IQR)</td>
<td>117 (28–124)</td>
<td>131 (36–170)</td>
<td>0.81</td>
<td>130 (28–185)</td>
<td>119 (38–115)</td>
<td>0.64</td>
</tr>
<tr>
<td>Randomized</td>
<td>31</td>
<td>30</td>
<td>0.82</td>
<td>34</td>
<td>27</td>
<td>0.03</td>
</tr>
<tr>
<td>Intention-to-treat</td>
<td>14</td>
<td>24</td>
<td>0.06</td>
<td>17</td>
<td>21</td>
<td>0.83</td>
</tr>
<tr>
<td>Double blinding</td>
<td>18</td>
<td>20</td>
<td>0.83</td>
<td>22</td>
<td>16</td>
<td>0.10</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>21</td>
<td>23</td>
<td>0.84</td>
<td>25</td>
<td>19</td>
<td>0.10</td>
</tr>
<tr>
<td>High overall quality</td>
<td>11</td>
<td>15</td>
<td>0.49</td>
<td>16</td>
<td>10</td>
<td>0.11</td>
</tr>
<tr>
<td>Bibliometric variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citations, median (IQR)</td>
<td>26 (21–48)</td>
<td>36 (3–48)</td>
<td>0.72</td>
<td>33 (2–55)</td>
<td>29 (3–46)</td>
<td>0.43</td>
</tr>
<tr>
<td>Recent publication</td>
<td>29</td>
<td>33</td>
<td>0.51</td>
<td>27</td>
<td>35</td>
<td>0.36</td>
</tr>
<tr>
<td>High-quality journal</td>
<td>17</td>
<td>21</td>
<td>0.53</td>
<td>21</td>
<td>17</td>
<td>0.21</td>
</tr>
</tbody>
</table>

IQR, interquartile range; ULT, ultimate truth; RTT, real-time truth.

### Outcome prediction

No statistically significant association between any variable and truth status was found in either truth dimensions following initial univariate cross-tabulation. Model building was continued due to the exploratory nature of the research question. Two-block forced-entry models were developed using assumed a priori variables of randomization, ITT analysis, blinding and sample size as block 1. Final best-fit models provided prediction rates (for true trials) of 59% (ULT) and 66% (RTT). Final model goodness-to-fit results indicated good prediction from the model to observed data: ULT $\chi^2 = 8.10$, $P = 0.42$; RTT $\chi^2 = 4.93$, $P = 0.77$. $R^2$ ranged from 0.08 to 0.10 for ULT and 0.13 and 0.18 for RTT. Thus, between about 8 and 10% for ULT and 13 and 18% for RTT of the outcome variance can be explained by the data. Beta values, $P$-values and Exph (odds ratios) for variables are presented in Tables 3 and 4. For ULT, only ITT was found to be a statistically significant predictor of outcome of truth status (false) of a trial ($P = 0.05$). For RTT, no statistically significant predictors of truth outcome appear in the final model, although publication year was near significance ($P = 0.06$). Exph for publication year is below one, so a more recent trial may reduce the odds of it being true by 66%. Odds ratios had wide confidence intervals crossing 1, thus generalization is poor. Analysis of residuals suggested no undue influence on the model from aberrant values (no outliers found with predicted probabilities, Cook’s distance or leverage values).

### Prediction of future trials

MCMC outcomes based on 46 true trials out of 93 total (49%) for ULT, and 44 true trials out of 93 (47%) for RTT are shown in Table 5.

The probability of at least 60% of all future trials being true ranged from <1 to 68%. A uniform prior represents no prior view or opinion as to whether or not the trial intervention could or should work. Uniform priors resulted in probabilities of at least 60% of future trials being true of 5% (ULT) and 2% (RTT). This probability was influenced by different priors reflecting either a sceptic opinion on the efficacy of the intervention: 0.001% both ULT and RTT; or an enthusiastic view: 66% for ULT and 68% for RTT.

### Discussion

This study hypothesized that although trial truth-status would be variable, better-quality trials are more likely to be true. Thus, quality variables could be used to predict trial truth status. Our findings claim that between 51 and 53% of published trials were false when judged against two scientific truth dimensions. The
only significantly different variable found between true and false trials was randomization in the RTT dimension.

The association of truth status and quality was neither consistent nor positive. When trial outcome was judged against the outcome of the most contemporary systematic review (ULT), only ITT analysis showed statistical significance as a predictor of truth status, but this was as a predictor for a false status (i.e. a trial was more likely to be false if ITT was used). When trial outcome was judged against an RTT dimension, no quality variable was shown to be a significant predictor. No significant associations were established with any bibliometric variable. Temporal trends suggested that trial truth status was not converging towards truth when truth is defined as the accumulation of trial findings, at any time point. Probability analysis demonstrated that the posterior estimates are dramatically influenced by prior beliefs, ranging from 0.001% for sceptic priors to 68% for enthusiastic priors. These data suggest that caution should be taken with inferences made without consideration of prior beliefs. It would take a high level of prior belief to use these data in suggesting that there was a reasonable chance of future trials being true.

<table>
<thead>
<tr>
<th>Name of Prior</th>
<th>Prior Distribution</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ULT</td>
<td>Uniform</td>
<td>0.49 (0.05)</td>
<td>0.08 (0.28)</td>
</tr>
<tr>
<td></td>
<td>Sceptic</td>
<td>0.14 (0.10)</td>
<td>0.001 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Enthusiastic</td>
<td>0.67 (0.18)</td>
<td>0.66 (0.19)</td>
</tr>
<tr>
<td>RTT</td>
<td>Uniform</td>
<td>0.47 (0.05)</td>
<td>0.04 (0.21)</td>
</tr>
<tr>
<td></td>
<td>Sceptic</td>
<td>0.14 (0.09)</td>
<td>0.001 (0.002)</td>
</tr>
<tr>
<td></td>
<td>Enthusiastic</td>
<td>0.67 (0.18)</td>
<td>0.68 (0.47)</td>
</tr>
</tbody>
</table>

ULT, ultimate truth; RTT, real-time truth.
by which other outcomes can be judged. They do, however, offer a
pragmatic method of deciding what ought to be accepted as truth by
a healthcare professional whose approach to clinical decision making
is embedded in a model of evidence-based practice. Clinical
guidelines are largely informed by systematic reviews of, ideally, randomized controlled trial-level outcomes. Therefore, con temporaneous guidelines or reviews are seen as a representation
of a real-world truth, at least in a normative sense. It could be
considered that these methods are as close as we can get to generating clinical, scientific truth. This aligns to the positivist principles
of clinical science and evidence-based practice. Although truth
variance of meta-analyses of reviews has been shown to be wide
[113], evidence-based practice continues to suggest that policy and
practice should be based on the outcomes of such reviews. Thus,
review outcomes should still be seen as scientific truthmakers.

The observed data are largely underdetermined by any single
theory. The findings might suggest error at any level of the
knowledge-generating process, i.e. trial level, systematic review-
ing or retrospective analysis such as the present study. The inten-
tion of quality variables is one of reducing bias in trial designs.

These data support existing findings which question the utility of
commonly accepted quality variables [8].

These findings contribute to contention and debate regarding the
nature and utility of controlled trials, as well as philosophical ex-
ploration of evidence-based practice [114]. In further agreement
with previous conclusions [1,2], such findings offer a fascinating
look at the process of science. We support the call for continued
inquiry into epistemic risk of health research, and concerns about
integration of evidence into practice [10–13].

Acknowledgements

The authors give thanks to James Coughlan, MCSP, MSc, University
of Nottingham, for classifying independently the truth status of
all reviews and trials.

References

1. Ioannidis, J. P. (2005) Contradicted and initially stronger effects
in highly cited clinical research. Journal of the American Medical Asso-
ciation, 294 (2), 218–228.

2. Ioannidis, J. P. (2005) Why most published research findings are
false. PLoS Medicine, 2 (8), e124.

3. Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C.,
Devereaux, P. J., Elbourne, D., Egger, M., Altman, D. G. & for the
CONSORT Group. CONSORT. (2010) Explanation and Elaboration: up-
dated guidelines for reporting parallel group randomised trial.
British Medical Journal, 340, c869.


5. Greenhalgh. T. (1997) How to read a paper – assessing the method-

Medical Journal, 112 (1081), 38–41.

controlled trials submitted to the ICS meeting in Heidelberg. Near-
ourology and Urodynamics, 24 (1), 21–24.

meta-analyses on the basis of quality scores should be abandoned.
Journal of Clinical Epidemiology, 59 (12), 1249–1256.

9. Morse, J. M. (2006) It is time to revise the Cochrane criteria. Quali-
tative Health Research, 16 (3), 315–317.

experiments. Studia Logica, 8, 215–246.

tives in Biology and Medicine, 52 (4), 500–517.


trolled trials in predicting effectiveness. Journal of Evaluation in
Clinical Practice, 16, 260–266.

WinBUGS – a Bayesian modelling framework: concepts, structure,


convergence of iterative simulations. Journal of Computational and

Graham, N., Haines, T., Bronfort, G. & Hoving, J. (2010) Manipu-
lation or mobilisation for neck pain: a Cochrane review. Manual
Therapy, 15 (4), 315–333.

[Objective criteria for the evaluation of chiropractic treatment of spon-
dylotic headache (author’s transl)]. Der Nervenarzt, 48 (5), 159–162.

lation for chronic neck pain. A double-blind controlled study. Spine,
7 (6), 532–535.

Pressure pain threshold evaluation of the effect of spinal manipula-
Manipulative and Physiological Therapeutics, 13 (1), 13–16.

effect of manipulation versus mobilization on pain and range of
motion in the cervical spine: a randomized controlled trial. Journal of
Manipulative and Physiological Therapeutics, 15 (9), 570–575.

Family Physician, 39, 1788–1790.

niques on pain and lateral flexion in neck pain patients: a pilot study.
ChiropracticTechnique, 8 (4), 155–162.

spinal manipulation in the treatment of cervicogenic headache.
Journal of Manipulative and Physiological Therapeutics, 20 (5),

tigating the effect of two manipulative approaches in the treatment
of mechanical neck pain: a pilot study. Journal of the Neuromusculosk-
etal System, 6 (1), 6–16.

clinical pilot trial comparing acupuncture, a nonsteroidal anti-
flammatory drug, and spinal manipulation. Journal of Manipula-
tive and Physiological Therapeutics, 22 (6), 376–381.

tigating the possible effect of the supine cervical rotatory
manipulation and the supine lateral break manipulation in the treat-
and Physiological Therapeutics, 23 (5), 324–331.

ized clinical trial on the relative effect of instrumental (MFMA)
versus manual (HVLA) manipulation in the treatment of cervical
spine dysfunction. Journal of Manipulative and Physiological Ther-
apeutics, 24 (4), 260–271.


