The authors' reply to the letter to the editor by Paul Dreyfuss et al.

Eugene J. Carragee · E. L. Hurwitz · I. Cheng · L. J. Carroll · M. Nordin · J. Guzman ·
P. Peloso · L. W. Holm · P. Côté · S. Hogg-Johnson · G. van der Velde · J. D. Cassidy ·
S. Haldeman · Secretariat of the Bone and Joint Decade 2000–2010
Task Force on Neck Pain and Its Associated Disorders

Received: 18 May 2008 / Accepted: 7 June 2008 / Published online: 4 July 2008
© Springer-Verlag 2008

The secretariat of the Bone and Joint Decade Task Force on Neck Pain and Associated Disorders welcomes the interest of Drs. Dreyfuss and Baker in our findings regarding radiofrequency neurotomy [1]. We are happy to elaborate why we found the two studies cited in their letter scientifically inadmissible to the best evidence synthesis.

This reply to the Letter to the Editor refers to the article
doi:[10.1007/s00586-008-702-1](https://doi.org/10.1007/s00586-008-702-1).

E. J. Carragee (✉)
Department of Orthopaedic Surgery,
Stanford University School of Medicine,
Stanford, CA, USA
e-mail: ecarragee@mac.com

E. J. Carragee
Orthopaedic Spine Center and Spinal Surgery Service,
Stanford University Hospital and Clinics,
Stanford, CA, USA

E. L. Hurwitz
Department of Public Health Sciences,
John A. Burns School of Medicine,
University of Hawaii at Mānoa, Hawaii, HI, USA

I. Cheng
Orthopaedic Surgery, Stanford University Medical Centre,
Stanford, CA, USA

L. J. Carroll
Department of Public Health Sciences,
Alberta Centre for Injury Control and Research,
School of Public Health, University of Alberta,
Alberta, AB, Canada

M. Nordin
Departments of Orthopaedics and Environmental medicine
and Program of Ergonomics and Biomechanics,
School of Medicine and Graduate School of Arts and Science,
New York University, New York, NY, USA

Firstly, we noted that both of these papers, the Lord et al. [8] and Govind et al. [5] publications report the findings from a single research center and have not been reproduced at any other location. We strongly believe the findings of those papers deserve close scrutiny by the spinal disorders community and serious reconsideration in the same manner as would be expected for any other

M. Nordin
Occupational and Industrial Orthopaedic Center (OIOC),
New York University Medical Center, New York, NY, USA

J. Guzman
Department of Medicine, University of British Columbia,
Victoria, BC, Canada

J. Guzman
Disability Prevention, Occupational Health and Safety Agency
for Healthcare in BC, Victoria, BC, Canada

P. Peloso
Endocrinology, Analgesia and Inflammation, Merck and Co.,
Rahway, New Jersey, NJ, USA

L. W. Holm
Division of Epidemiology, Institute of Environmental Medicine,
Karolinska Institutet, Stockholm, Sweden

P. Côté
Department of Public Health Sciences, University of Toronto,
Toronto, ON, Canada

P. Côté
University Health Network Rehabilitation Solutions,
Toronto Western Hospital, Institute for Work and Health,
Toronto, ON, Canada

S. Hogg-Johnson
Institute for Work and Health, Toronto, ON, Canada

contemplated treatment approach that was based on a very limited or poorly validated research.

The methodological problems with the Lord et al. [8] study are itemized in our best-evidence synthesis [1]. At the heart of the methodological flaws are the issues of (1) unbalanced randomization in a very small study population (only 12 subjects receiving the intervention) and (2) the extension of outcomes analysis beyond the time of blinding and validated outcomes assessment (which by design was limited to 3 months or less).

Regarding the unbalanced randomization, this methodological problem was severe and the number of subjects was so small that no statistical adjustment for confounding variables can be made. Lord et al. reported that twice as many subjects in the sham group (33%) had painful facet joints that were not treated. If the authors are correct, one would suspect some continued neck pain in 33% of the sham group on this basis alone. That is before the intervention was even done, the sham group may have had only eight potential subjects who could report complete neck pain relief (the success criterion) with four cases having untreated neck problems necessarily falling within the failed treatment category. To have a “controlled” study with only a small number of subjects, randomization not only should be performed, it needs to be successful in providing similar populations of patients in the treatment and controlled groups.

Along the same line of unbalanced randomization, 83% of the sham group had litigation related to the neck pain syndrome being treated (vs. 33% of the RF group). Drs. Dreyfuss and Baker cite the work of the senior author of the Lord et al. [8] and a study of Sapir and Gorup [9], to contend that litigation is clinically irrelevant in the context of chronic whiplash disorders. To most active clinicians, this proposition does not pass “the chuckle test,” nor does

it comport with the literature in general [2, 3, 6]. In fact, the study by Sapir and Gorup cited by Drs. Baker and Dreyfuss actually reported worse outcomes in active litigation subjects undergoing RF neurotomy (at 2 week 15% more non-litigation subjects achieved >70% pain reduction after RF, and at 1 year a higher pain VAS of 4.0 in the litigation group versus 2.9 (non-litigation), $P = 0.05$).

More telling is the study by Dr. Dreyfuss himself, on RF for facet pain [4]. In that study Dr. Dreyfuss excluded subjects with “compensable disability or work injury, [or] ongoing litigation” because these factors were “comorbid or psychosocial factors...likely to confound the results.” We completely agree with Dr. Dreyfuss’ own assessment: compensation disputes and litigation are very likely to confound a study’s results. So when the Lord et al. study reported far more of the sham group had active litigation issues there is little confidence that outcomes differences in favor of the RF procedure can be attributable to the intervention alone.

Next, there is the issue of loss of blinding due to documented sensory changes in nearly half the RF group (none in the sham group) soon after the procedure and additional selective un-blinding at 3 months. Our group had, again, little confidence that either the evaluating surgeon or the subjects were actually blinded to which treatment was received in these subjects and expect some reporting bias is possible for that reason.

Finally, after the blinding protocol had been broken at 3 months, subjects were then followed by the un-validated (and likely inaccurate) method of being directed to call back when their pain exceeded 50% of their pre-procedure intensity. Lord et al. counted any subject they did not hear from as a success! This assumption is, charitably stated, dubious. Therefore, although a Kaplan–Meier and Cox Proportional Hazard analysis may have been reasonable if

S. Hogg-Johnson
Department of Public Health Sciences, University of Toronto,
Toronto, ON, Canada

G. van der Velde
Department of Health Policy, Management and Evaluation,
University of Toronto, Toronto, ON, Canada

G. van der Velde
Institute for Work and Health, Toronto, ON, Canada

G. van der Velde
University Health Network Rehabilitation Solutions,
Toronto Western Hospital, Toronto, ON, Canada

G. van der Velde
Division of Health Care Outcomes and Research,
Toronto Western Research Institute, Toronto, ON, Canada

J. D. Cassidy
Department of Public Health Sciences, University of Toronto,
Toronto, ON, Canada

J. D. Cassidy
Division of Health Care Outcomes and Research,
Toronto Western Research Institute, University Health
Network, Toronto, ON, Canada

S. Haldeman
Department of Neurology, University of California,
Irvine, IR, USA

S. Haldeman
Department of Epidemiology, School of Public Health,
University of California, Los Angeles, CA, USA

the subjects had been followed with standard metrics every month or so, this was not the case. The most valid data appear to be that collected before the patient-initiated “call back assessment” method was begun (3 months). For those few data points the Fisher Exact test is an appropriately applied analysis for small subject studies. Clearly, there was no statistical difference in outcomes at 3 weeks or at 3 months—despite the randomization imbalances strongly favoring the RF group.

The Govind et al. [5] study, again by the same research group, on RF neurotomy for suspected cervicogenic headache was also felt to be scientifically inadmissible by the Task Force. While Govind et al., do report “extraordinary” results, they unfortunately also report less than extraordinary methods: this was an unblinded, uncontrolled, non-randomized case-series report. No independent examiner, nor standard pain metrics, nor functional questionnaires were used to assess outcome. This case-series was promptly criticized for these methodological flaws in the same journal [7]. Clearly, exploratory studies such as Govind et al. are important and are needed to direct more definitive clinical research. They are, however, the least substantial of all forms of clinical evidence, and a poor basis for advocating general clinical application of invasive treatments.

Both Dr. Dreyfuss and Dr. Baker are careful clinicians and scientific thinkers, whom we respect immensely. In fact we expected, at the onset of our best-evidence synthesis, to find much more substantial evidence supporting RF in this area. We, like Drs. Dreyfuss and Baker, believed that RF neurotomy might have had good data supporting treatment for neck pain in appropriately selected patients. However, we did not find this and, frankly, if these two highly-flawed papers are the “best evidence” for the use of RF neurotomy in neck pain syndromes, this is sad news indeed—and well out of step

with efforts to increase standards for evidence-based clinical care. After careful review of this area of treatment, it is clear that the most basic studies still need to be performed. If these are highly effective treatments, it should be easily proven in a properly designed trial, as we outlined in our research recommendations [1], and testing the hypothesis should be a priority for those interested in this type of intervention.

References

1. Carragee EJ, Hurwitz EL, Cheng I et al (2008) Treatment of neck pain: injections and surgical interventions: results of the Bone And Joint Decade 2000–2010 Task Force On Neck Pain And Its Associated Disorders. *Eur Spine J* 17(Suppl.1):S153–S169
2. Carroll LJ, Holm LW, Hogg-Johnson S et al (2008) Bone and Joint Decade 2000–2010 Task Force On Neck Pain And Its Associated Disorders. Course and prognostic factors for neck pain in whiplash-associated disorders (WAD): results of the Bone and Joint Decade 2000–2010 Task Force On Neck Pain And Its Associated Disorders. *Spine* 33(4 Suppl):S83–S92
3. Cassidy JD, Carroll LJ, Cote R et al (2000) Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash disorder. *NEJM* 342:1179–1186
4. Dreyfuss P, Halbrook B, Pauza K et al (2000) Efficacy and validity of radiofrequency neurotomy for chronic lumbar zygapophysial joint pain. *Spine* 25(10):1270–1277
5. Govind J, King W, Bailey B, Bogduk N (2003) Radiofrequency neurotomy for the treatment of third occipital headache. *J Neurol Neurosurg Psychiatry* 74:88–93
6. Kasch H et al (2001) Handicap after acute whiplash injury: a 1-year prospective study of risk factors. *Neurology* 56:1637–1643
7. Kwan O, Friel J (2003) Radiofrequency neurotomy. *J Neurol Neurosurg Psychiatry* 74(8):1164–1165
8. Lord SM, Barnsley L, Wallis BJ et al (1996) Percutaneous radiofrequency neurotomy for chronic zygapophyseal-joint pain. *N Engl J Med* 335:1721–1726
9. Sapir DA, Gorup JM (2001) Radiofrequency medial branch neurotomy in litigant and nonlitigant patients with cervical whiplash. *Spine* 26:E268–E273