Reviewer’s report

Title: Assessment of nerve involvement in the lumbar spine: association between magnetic resonance imaging, physical examination and pain drawing findings

Version: 3 Date: 16 April 2009

Reviewer: Charlotte Leboeuf-Yde

Reviewer’s report:

I have previously reviewed this manuscript. I then commented (only) on the major issue of how the data were reported because I considered this to be so important that the rest was irrelevant. The authors seem to have accepted a few of the reviewers’ comments. The manuscript is well written and easy to read and the tables are clear, and it is good that some of them have been simplified with the possibility to get access to the more detailed data. However, in my opinion, the manuscript still suffers from the same major weakness as before.

In this review I shall restate this issue and have added my other concerns that I did not bother with last time, because I assumed that the manuscript would be subjected to a major re-write. However, as I can see that the other reviewers have not commented on the same issue as I have, it is possible that they will accept the manuscript as is (which is fair enough). However, in that case I think it is important that all my comments are taken into consideration.

In short: This manuscript suffers from three major weaknesses: First, the study sample is far too small for all these analyses. Second, the rationale behind the analyses is too simplistic because it is not clinically relevant to study single clinical findings taken out of context in relation to the MRI-findings. In stead it necessary to make a clinical judgement of all clinical findings in relation to whether there is a nerve-involvement or not. Third, your interpretation of the findings does not take into account the fact that it when the discal prolapse is gone, it is fairly common with long-term sequelae such diminished altered sensitivity in the previously acutely affected areas. Therefore it is logical that in chronic LBP patients with a history of nerve root affection, there would be clinical findings that cannot be accounted for with MRI-inspection of nerve root interference.

Please see detailed comments below.

Major compulsory revision.

1. Regarding the (in my opinion) lack of clinical relevance in your analytic strategy: You decided to cross-tabulate individual detailed nerve-root-related MRI variables with single specific clinical findings. Your conclusion is that there are fewer MRI findings than clinical findings. You also note that when using the clinical findings as gold standard, the MRI examination is not particularly clinically useful, because the clinical findings are not matched by the MRI findings.
However, you must ask yourselves this question: Is it likely that clinicians would examine a patient, find that there is, for example, reduced sensitivity at one specific spinal level or diminished muscle strength, and make a clinical decision thereupon? No, of course not! I shall therefore insist that you need to include not individual clinical examination findings but the clinical conclusion as to whether there is likely nerve root involvement or not (of course at each individual spinal level). If you are keen on including the detailed information so that the reader can follow your thinking and get the background information, that is OK with me, but only in addition to the (yet to define) clinical conclusion variables. This clinical conclusion should be made independently of the MRI information.

2. There are still far too many statistical tests in this little study of 61 persons. At least, discuss this as a weakness in the discussion section or else adjust for it somehow, by demanding a lower p-value than 0.05.

3. Reliability is an important issue. It being so poor in the lower spinal levels for MRI readings merits to be discussed under limitations in the discussion section. It is not sufficient to just accept the readings from the radiologist who had the most positive findings. They may be many but they may still be incorrect.

4. The discussion very briefly touches the issue of reliability for the clinical findings, referring to two articles that most readers would not have the energy to download or order home. It might be a better idea to devote a paragraph or two in the methods section to the validity of the three types of variables (MRI, clinical examination, pain drawing- own present findings and past findings) and provide a proper presentation of their credibility. Any weaknesses in this area should be taken under consideration when interpreting the data.

5. In relation to your analytic strategy, it is not always possible to understand what, for you, constitutes a normal finding in relation to the MRI assessment because your “additional definitions” in the methods section do not define normality. For example, what is normal if “slight” is defined as <50% decreased disc height? In other words, the “additional definitions not noted in the MRI protocol” need to be gone through with a fine comb. Also, why are they “additional”? Are they post hoc definitions developed in the course of the data analysis because the other variables were found to be lacking?

6. Was the kappa analysis for the MRI variables applied on the detailed findings or on overall findings? This is not clear from the methods section and it would have consequences on the possibility to get acceptable agreement.

7. I do not understand your statistical methods. Are you looking for associations or agreement? Is the McNemar’s test really the method of choice and is it interpreted correctly? If you have not used a professional statistician on this issue, please make sure that you do check with such a person. I am confused by your tables, such as your Table 3 and Ad. Table 3. In the heading you write “Association between...” but in the table itself you write “Agreement” in your columns relating to sensitivity to touch, to pain etc. Is this about associations, which McNemar’s test should be able to handle, or is it about agreement, when
your statistical test (I believe) should be handled differently? This requires at least clarification.

8. Because your nerve disturbance findings through the clinical examination cannot be considered a gold standard for a concomitant pathological mechanical nerve interference, it could be argued that it is incorrect to calculate sensitivity and specificity. Associations between the two sets of findings would be a more appropriate way of considering this issue perhaps.

9. In the result section, you need to separate the reliability results from the descriptive data and the cross-tabulations.

10. Finally, two comments, one important one less so, perhaps. It has been shown in several studies that a relatively large proportion of patients with previous discal hernia experience long-term sequelae thereof. It has also been shown that discal prolapses in at least half of the time will diminish considerably without surgical intervention (See TS Jensen and Hanne Albert as one of the co-authors), and of course after surgery they will disappear as well. As you are working with chronic patients, many of these are likely to have had a discal prolapse which has now receded. It is also known that people who recovered from sciatica often experience sensory sequelae. This means that it is to be expected that there are more clinical findings of nerve injuries than MRI findings of acute nerve interference. You need to devote a considerable amount of space in your discussion section to relate to this fact, including bringing references from studies showing this to be the case.

11. My minor point on this issue is that therefore it is not a good idea to recommend yet another study with a heterogeneous patient material. On the contrary, if you want to study this further, it should be on a homogeneous patient group consisting of acutely ill patients, to check if it is not true that when the discal hernia is “fresh” its findings will correspond with a logical clinical picture. If you want to do this study on a heterogeneous population, it would need to be very large, to make it possible to control for duration of pain and possibly age and sex.

Minor issues

1. I am not sure that results will become less “reliable” just because there are few findings. The statistical tests can perhaps not be performed and too few estimates in each cell may make the data weak (but “reliable”?).

2. I noted a weird conclusion, based on refs 19-22 on page 15 in your discussion section. “Non-specific pain per se may be more detrimental than pain due to a known cause and may contribute to the poor outcome of low back pain treatment”. I know some of these references and I cannot understand how you conclude what you did in this phrase. These studies have not looked at the issue you mention here. You need to check that out or remove the references, as your statement as such makes perfectly sense and does not need to be supported by any experimental data of any sort.
3. Background, line 6, word missing after “originating”.

4. Further down on the same page you seem to consider a clinical examination to be “objective”. Clinical examination is fairly subjective, depending on the subjective reporting by the patient and the subjective interpretation of this reporting by the clinicians. Please explain how you can state it to be subjective or rectify.

5. Methods, first para, the study started in September 2004 but it was completed when?

6. P. 7. I do not understand the sentence “Each radiologist made independent assessments that were repeated before and after reading the patient’s history from the referring physician” in your first para.

7. A reference for your McNemar’s test is needed in the analytical methods, if you really will keep this test (after discussions with statistician).

8. Results section under prevalence of findings. What about the medulla cord signal and bone protuberance? How were they different from the others?

9. P.10 under “Pain drawing findings of nerve involvement”. The first sentence of that paragraph does not stand on its own. What was found in 95% of patients?

10. P.13. Your statement that the sensitivity that increased significantly from 16% to 29% surely lacks clinical interest. Why mention it at all?

11. In the discussion section, p.14 second line you speak of associations but where were these reported?

12. In the tables the word “dichotomised” has been written “dikotomised”, whereas it is correct throughout the text.