Author's response to reviews

Title: Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain

Authors:

- Francisco M Kovacs (kovacs@kovacs.org)
- Victor Abraira (victor.abraira@hrc.es)
- Ana Royuela (victor.abraira@hrc.es)
- Josep Corcoll (jcorcoll@dgpif.caib.es)
- Luis Alegre (lalegre@ibsalut.caib.es)
- Miquel Tomas (miqueltomas@ibsalut.caib.es)
- Maria Antonia Mir (mamir@ibsalut.caib.es)
- Alejandro Cano (victor.abraira@hrc.es)
- Alfonso Muriel (alfonso.muriel@hrc.es)
- Javier Zamora (javier.zamora@hrc.es)
- Maria Teresa Gil del Real (mtgildelreal@kovacs.org)
- Mario Gestoso (mgestoso@kovacs.org)
- Nicole Mufraggi (kovacs@kovacs.org)

Version: 6 Date: 18 February 2008

Author's response to reviews: see over
In addition to the replies to the reviewer’s comments, the authors wish to point out that two paragraphs have been added to the manuscript, in order to update the bibliography with a paper on this very subject that has been coauthored by the reviewer and very recently published. Those paragraphs are copied at the end of this document.

Author’s replies to the additional comments by the reviewer are as follows:

Major compulsory revisions

1.1 Definitions MDC and MCIC, and MCS and ROC.

1.1.a: In the abstract, the definitions of MCIC and MDC are well adapted now. In the introduction, line 4 the term “variation” is still used instead of “change” and “that measure a symptom” can be deleted. So the definition becomes: “MCIC is defined as the minimal change in score that is meaningful for patients”.

“Variation” was used to avoid using “change” twice (in the name –MCIC- and its definition). However, in accordance with the reviewer’s comment, the definition now reads (first paragraph under Introduction):

Minimal detectable change (MDC) is defined as the minimal change that falls outside the measurement error in the score of an instrument used to measure a symptom. Minimal clinically important change (MCIC) is defined as the minimal change in the score that is meaningful for patients.1-12

1.1.b Page 5, line 9: MDC and MCIC are different constructs and MCS and ROC are both methods to assess MCIC.

The authors agree. In fact, that is exactly what is stated in the manuscript:

- First paragraph under Introduction:

Minimal detectable change (MDC) is defined as the minimal change that falls outside the measurement error in the score of an instrument used to measure a symptom. Minimal clinically important change (MCIC) is defined as the minimal change in the score that is meaningful for patients.1-12 Different approaches can be used to determine MCIC. One is to estimate the mean change in score in patients who actually report to have improved (referred to as “mean change score”, or MCS). Another approach is to use receiver operating characteristics curves (ROC) to define the cutoff point that best discriminates between patients reporting or denying any improvement.
The MDC and MCIC values in the PI-NRS for neck and referred pain were estimated for the follow-up period of 3 months.

The Minimal Detectable Change (MDC) was calculated as $1.96 \times \sqrt{2} \times \text{SEM}$. The standard error of measurement (SEM) was estimated by taking the square root of the within subject variance (consisting of variance between measures plus the residual variance on a two-way ANOVA random effects model) of patients categorized as “unchanged” by external criterion. The 95% confidence interval was calculated using the chi-square distribution. The MDC can be interpreted as the magnitude of change below which there is more than a 95% chance that no real change has occurred.

The following methods were used to estimate the MCIC:

1. Mean Change Score (MCS): Mean change of PI-NRS in patients who scored “2” (“improved”) on the external criterion. The changes of scores PI-NRS were calculated by subtracting the final values from the baseline values, so that positive scores correspond to improvement.

2. Optimal cutoff point (ROC): Considering the PI-NRS change as a diagnostic test for discriminating between improved and not improved patients, and the external criterion as a gold standard, a ROC curve was developed describing the performance of changes in the corresponding scale to detect improvement. The optimal cutoff point was estimated by the point that maximizes the sum of specificity and sensitivity.

### 1.2 Numbers

1.2.1

The numbers appear to be very small. In my opinion it is not justified to calculate MDC values for samples < 20 persons. So I should not present these values for the subgroups. You might present MCS and ROC values for the subgroups though.

**In accordance with the reviewer's comment, we have maintained MCS and ROC values for subgroups, and have omitted MDC values for subgroups in which the number of subjects was too small. The manuscript has been modified accordingly and it now reads:**

- Table 3:

<table>
<thead>
<tr>
<th>Measurement of change</th>
<th>Value (all 658 patients)</th>
<th>Baseline pain severity (PI-NRS)</th>
<th>Chronicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lowest tertile*</td>
<td>Subacute</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Highest tertile*</td>
<td>Chronic</td>
</tr>
</tbody>
</table>
**PI-NRS**: Pain intensity numerical rating scale.
**MDC**: Minimal detectable change. **MCS**: Mean change score. **ROC**: Optimal cutoff point on the ROC curve.

*: Lowest and highest tertiles do not represent the same number of patients because of repeated scores for pain severity.

1.2.1: ****: The low number of patients denying any change in this sample made it impossible to reliably estimate the potential effect of baseline pain severity and chronicity status on MDC values.

### Table 4:

Table 4. MDC, MCS and ROC values for neck pain and referred pain only in patients with referred pain at baseline, and differences depending on baseline pain severity and chronicity

<table>
<thead>
<tr>
<th>Measurement of change</th>
<th>Value (all 487 patients with referred pain at baseline)</th>
<th>Baseline pain severity (PI-NRS)</th>
<th>Chronicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest tertile* (&lt; 6 points)</td>
<td>Highest tertile* (≥ 8 points)</td>
<td>Subacute (&lt;90 days)</td>
</tr>
<tr>
<td><strong>Neck pain (PI-NRS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDC; Value (95% CI)**</td>
<td>4.0 (3.4-5.0)</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td>n</td>
<td>48</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>MCS (SD)</td>
<td>4.1 (2.3)</td>
<td>2.6 (1.4)</td>
<td>4.9 (2.3)</td>
</tr>
<tr>
<td>n</td>
<td>395</td>
<td>78</td>
<td>209</td>
</tr>
<tr>
<td>ROC curve</td>
<td></td>
<td>0.92 (0.87-0.96)</td>
<td>0.94 (0.85-1.00)</td>
</tr>
<tr>
<td>Area (IC 95%)</td>
<td></td>
<td>0.94 (0.86-0.97)</td>
<td>0.95 (0.91-0.99)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.93</td>
<td>0.89</td>
<td>0.96</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.83</td>
<td>0.93</td>
<td>0.80</td>
</tr>
<tr>
<td>ROC</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>N</td>
<td>658</td>
<td>164</td>
<td>321</td>
</tr>
<tr>
<td><strong>Pain referred to the arm (PI-NRS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDC; Value (95% CI)**</td>
<td>6.2 (5.0-8.2)</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>n</td>
<td>35</td>
<td>2.4 (1.7)</td>
<td>3.7 (3.0)</td>
</tr>
<tr>
<td>MCS (SD)</td>
<td>4.1 (2.7)</td>
<td>96</td>
<td>127</td>
</tr>
<tr>
<td>n</td>
<td>304</td>
<td>114</td>
<td>249</td>
</tr>
<tr>
<td>ROC curve</td>
<td></td>
<td>0.81 (0.73-0.89)</td>
<td>0.80 (0.65-0.96)</td>
</tr>
<tr>
<td>Area (IC 95%)</td>
<td></td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.71</td>
<td>0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>Specificity</td>
<td>1.5</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>ROC</td>
<td>479</td>
<td>159</td>
<td>201</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PI-NRS**: Pain intensity numerical rating scale
**MDC**: Minimal detectable change. **MCS**: Mean change score. **ROC**: Optimal cutoff point on the ROC curve.

*: Lowest and highest tertiles do not represent the same number of patients because of repeated scores for pain severity.
1.2.1: **The low number of patients denying any change in this sample made it impossible to reliably estimate the potential effect of baseline pain severity and chronicity status on MDC values.**

- **Fifth paragraph under Discussion:**

  Although the low number of patients in this sample denying any change made it impossible to reliably estimate the potential effect of baseline pain severity and chronicity status on MDC values (Tables 2-4), results from this study also show that MDC for neck pain is similar to MCS, both in all the patients and in the subset of those complaining from referred pain. Although MDC and MCS are different concepts and different methods are used to calculate their values, their size is very similar. This finding is consistent with results from previous studies, both in neck and low back pain patients.\(^{8,9,16}\) It might be interpreted that the limit of what constitutes a “relevant” change is similar whether it derives from the scores of patients who report improvement or deny it. However, as opposed to ROC, those calculations do not take into account false positives and false negatives.

1.2.2

You say in this response that you don’t have normal distributions. However the assumption to calculate \(1.96 \times \sqrt{2} \times \text{SEM}\) is that the distribution is normal. So MDC cannot be calculated in case of skewed distributions.

*The reviewer’s comment is correct. Normality is an assumption to calculate MDC and, as we stated in our last response, samples for some of our subgroups depart from normality. However, given that there is no other method to compute MDC, we prefer to maintain these estimations in all subgroups for completeness and in order to allow comparison of our results with other existing evidence.*

*In accordance with the reviewer’s comment, the manuscript has been modified to further clarify this issue, and it now reads (middle of third paragraph under Discussion):*

  MDC computation relies on the assumption of data being drawn from a normally distributed population. Although some of the subgroups defined by chronicity or pain severity were skewed, results for all of them are presented for completeness and to allow comparisons with results from similar studies (Tables 3 and 4).\(^{8,9,16}\)

1.2.3

I wonder how you calculated the MDC for referred pain, where the improvement is 1.6 points and the SD = 2.7 (n=35). How do you arrive at a SEM value?

*SEM has been calculated following the standard methods, as described in the second paragraph under Analysis:*

  The Minimal Detectable Change (MDC) was calculated as \(1.96 \times \sqrt{2} \times \text{SEM}\).\(^{8,12}\) The standard error of measurement (SEM) was estimated by taking the square root
of the within subject variance (consisting of variance between measures plus the residual variance on a two-way ANOVA random effects model) of patients categorized as “unchanged” by external criterion.

It may be that the reviewer has mistakenly used the data she mentions (SD = 2.7 and n=35) to compute “Standard Error of the Mean” instead of “Standard Error of the Measurement”. This might have misled her to conclude that we have erroneously calculated SEM, when actually there is no error there.

1.2.4 It is misleading to calculate a 95% CI around the MDC as the MDC is already the upper 95% limit of a distribution. In fact, the MDC value is not a confidence interval; it is a point estimator of the upper limit of the error of measurement, and it is computed using a point estimation of a variance component. Like any other point estimation, it is susceptible to be estimated also as a 95% confidence interval. The authors strongly feel that it is important to give an idea of the precision of the estimation of MDC, such as the 95% CI provides.

1.2.5 Discussion, par 2. No interpretation is give for the similarity between MCS and MDC. What does that mean. And the authors should comment on the dissimilarity between ROC and MCS, as it are both assessments of MCIC. In this regard, the text in the fourth par of the discussion (MCS and ROC represent different constructs) is not convincing.

The authors don’t explain why different values are found and which are more valid. The text in the paragraph comes down to saying you can use whatever value you want. Taking into account that MCS and ROC correspond to different concepts, are estimated using different methods and include scores from different subsets of patients (i.e., only those who have improved vs. also from patients denying any improvement), it is not strange that their values are different.

The authors feel unable to state which method “is more valid”, since it depends on the objective they are used for (i.e., valid for what).

The authors feel that summarizing the text in the paragraph as meaning that “you can use whatever value you want” is a potentially misleading oversimplification. Therefore, they have further clarified the text in the updated version of the manuscript.

In accordance with the reviewer’s comment, the manuscript now reads (fifth and sixth paragraphs under Discussion):

Although the low number of patients in this sample denying any change made it impossible to reliably estimate the potential effect of baseline pain severity and
chronicity status on MDC values (Tables 2-4), results from this study also show that MDC for neck pain is similar to MCS, both in all the patients and in the subset of those complaining from referred pain. Although MDC and MCS are different concepts and different methods are used to calculate their values, their size is very similar. This finding is consistent with results from previous studies, both in neck and low back pain patients.\(^8\),\(^9\),\(^16\). It might be interpreted that the limit of what constitutes a “relevant” change is similar whether it derives from the scores of patients who report improvement or deny it. However, as opposed to ROC, those calculations do not take into account false positives and false negatives.

MCS and ROC represent different constructs and the methods for calculating them differ, so it is not surprising that their values are different, with ROC being consistently smaller than MCS (Tables 3 and 4).\(^8\),\(^9\) Although it is up to researchers or clinicians to decide whether MCS or ROC are more suitable to define MCIC in their specific circumstances, the consistency of ROC and MCS values across studies may help them to use these results in practice (Tables 3 and 4).\(^8\),\(^9\),\(^16\) The upside of using the MCS value instead of ROC is that patients with scores showing an improvement above its value have a 95% chance of having improved meaningfully. However, in general, ROC might be more suitable, since scores from patients both reporting and denying improvement are used to calculate it, and it tends to weigh equally false-positive and false-negative misclassifications.\(^16\) As it has been suggested, the choice between the two methods may also depend on the type of intervention or the clinical consequences of being a “false positive” or “false negative.”\(^30\) For instance, some researchers may prefer to anticipate a difference generally corresponding to ROC (e.g. 1.5 PI-NRS points) for sample calculations in clinical trials vs. placebo, since ROC represents “the cut-off point that best discriminates between those patients feeling and not feeling that they have improved” and, since its size is smaller than MCS, it leads to larger samples. On the contrary, some clinicians may prefer to disregard differences smaller than MCS (e.g., 4 PI-NRS points) when they have to select among treatments with different safety profiles or side effects for a given patient, since that value represents “the mean change above which most patients would feel they have improved”.

1.2.6 Problems with the numbers and data in the Table
The specificity = 1.00 (Table 4, subgroup sub acute): in that case the ROC should be larger than the MDC value. The finding that the MDC (=3.1) is larger than the ROC value (=0.5) may be due to small numbers (and should not be presented in that case), but the value of 1.5 for the ROC in the total group of neck pain with referred pain (Table 4, first column) cannot be right.

In this example, it is impossible that the ROC value is 1.5, when the mean value of the stable group is already larger (i.e. 1.6 see Table 2). Please explain what is going on here. Drawing an anchor based MIC distribution\(^1\) is an adequate method to get more insight into the data and explain the differences between the ROC and MCS values.

1. De Vet et al. Minimally important change determined by a visual method
integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007; 16(1):131-42.

In accordance with reviewer’s comment 1.2.1, we have maintained ROC and MCS for all subgroups and deleted MDC values for subgroups with a too small sample size, as in the instance raised here (please, see point 1.2.1 in this document, and tables 3 and 4 in the updated version of the manuscript).

The reviewer states that it is impossible to reach one of the numerical results from this study, namely for ROC value to be 1.5 when the mean value of the stable group is 1.6. However, those are the actual data. The authors are willing to provide the editor with the raw data, to make it possible to repeat the corresponding calculations by an independent party. The authors are fully aware that this is an unusual proposal, but they feel that the reviewer’s statement is also unusual, and they feel that it would be the best way to show the correctness of these results.

1.3 Chi square distribution.
OK, it is possible that for small numbers the distribution is called X2 distribution.

The authors agree: the X2 distribution can be called the X2 distribution, even if with large numbers that distribution may approach the normal one.

1.4; 1.5; 1.6
OK

1.7 Interpretation of MCIC values.
The authors focus more on consistency with other studies than on consistency within their own study. As said under point 1.2 the content of the fourth par of the discussion is not convincing. Can a researcher choose to use either 1.5 or 4.0 for power calculations and for interpretation of trials?

In accordance with the reviewer’s comment, the manuscript has been modified to read (fifth and sixth paragraphs under Discussion):

Although the low number of patients in this sample denying any change made it impossible to reliably estimate the potential effect of baseline pain severity and chronicity status on MDC values (Tables 2-4), results from this study also show that MDC for neck pain is similar to MCS, both in all the patients and in the subset of those complaining from referred pain. Although MDC and MCS are different concepts and different methods are used to calculate their values, their size is very similar. This finding is consistent with results from previous studies, both in neck and low back pain patients. It might be interpreted that the limit of what constitutes a “relevant” change is similar whether it derives from the scores of patients who report improvement or deny it. However, as opposed to ROC, those calculations do not take into account false positives and false negatives.
MCS and ROC represent different constructs and the methods for calculating them differ, so it is not surprising that their values are different, with ROC being consistently smaller than MCS (Tables 3 and 4).\textsuperscript{8,9} Although it is up to researchers or clinicians to decide whether MCS or ROC are more suitable to define MCIC in their specific circumstances, the consistency of ROC and MCS values across studies may help them to use these results in practice (Tables 3 and 4).\textsuperscript{8,9,16} The upside of using the MCS value instead of ROC is that patients with scores showing an improvement above its value have a 95% chance of having improved meaningfully. However, in general, ROC might be more suitable, since scores from patients both reporting and denying improvement are used to calculate it, and it tends to weigh equally false-positive and false-negative misclassifications.\textsuperscript{16} As has been suggested, the choice between the 2 methods may also depend on the type of intervention or the clinical consequences of being a “false positive” or “false negative.”\textsuperscript{30} For instance, some researchers may prefer to anticipate a difference generally corresponding to ROC (e.g., 1.5 PI-NRS points) for sample calculations in clinical trials vs. placebo, since ROC represents “the cut-off point that best discriminates between those patients feeling and not feeling that they have improved” and, since its size is smaller than MCS, it leads to larger samples. On the contrary, some clinicians may prefer to disregard differences smaller than MCS (e.g., 4 PI-NRS points) when they have to select among treatments with different safety profiles or side effects for a given patient, since that value represents “the mean change above which most patients would feel they have improved”.

2.1 to 2.10 OK

2.11 Discussion on influence of intervention
Page 11, line 3 The reference of Farrar refers to different pain conditions and does not say so much about interventions.

In fact, that is precisely what the text says (middle of ninth paragraph under Discussion):

In fact, MCIC seem to be consistent even across different chronic pain conditions.\textsuperscript{11}

In addition, in accordance with the reviewer’s comment, the manuscript now reads (ninth paragraph under Discussion):

Mean duration of pain when patients entered this study was over 540 days (Table 1). During that period, they had all received many forms of treatment and many still received them during the study.\textsuperscript{22} Since data being analyzed in this study derive from post-marketing surveillance of neuroreflexotherapy, all of them received that specific form of treatment.\textsuperscript{22-27} No study has assessed the potential influence of any specific form of treatment on MDC or MCIC and many studies include patients receiving heterogeneous treatments, since they are participating in randomized controlled trials.\textsuperscript{8,16} In fact, MDC and MCIC calculations rely on patients’ self-assessment of their own evolution and scores from instruments used to assess evolution of symptoms, no matter what treatments are potentially influencing that evolution. In fact, MCIC seem to be consistent even across different chronic pain conditions.\textsuperscript{11} Therefore, the generalizability of results from
this study are not affected by the fact that these results derive from the post-marketing surveillance of a particular form of treatment. The consistency of results from this study with those from previous reports on neck pain and low back pain patients further supports their generalizability.

2.12 Next paragraph: first line: “On the contrary” can be deleted.

In accordance with the reviewer’s comment, this major compulsory revision has been undertaken, and “on the contrary” has been deleted. Therefore, the updated version of the manuscript now reads (beginning of tenth paragraph under Discussion):

In fact, using post-marketing surveillance methods in a National Health Service to assess MCIC has a number of advantages.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests

In addition to these replies to the reviewer’s comments, the paragraphs that have been added to update the bibliography with the paper coauthored by the reviewer that has been very recently published, are as follows:

Fourth paragraph under Introduction:

In patients with nonspecific neck pain (NP), MDC and MCIC have been explored for disability and quality of life.\(^{13-15}\) MDC and MCIC for pain severity have also been recently explored.\(^{16}\) However, MDC and MCIC for pain were only explored in a small sample of patients participating in a randomized controlled trial, it is unknown whether or not they had referred pain, and approximately 75% of them were acute patients,\(^{16}\) while subacute and chronic patients represent the major part of the social, clinical and economical burden associated with spinal disorders.\(^{17}\) MCIC for neck pain might be different between acute and chronic patients, between patients with and without referred pain, and between patients seen in routine clinical practice and those included in a randomized controlled trial in which expectations, and other unspecific effects (such as Hawthorne or placebo) might also influence them.

Therefore, the primary objective of this study was ….
In spite of differences in methods commented on in the Introduction section, present results are consistent with those from the only previous study in which MDC and ROC were explored for neck pain.\textsuperscript{16} In that study, MDC for neck pain was 4.3 PI-NRS points, and ROC was 2.5.\textsuperscript{16} In the current study, those values were 4.0 and 1.5.