

Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach

Henrica C. W. de Vet^{1,5}, Raymond W. J. G. Ostelo^{1,2}, Caroline B. Terwee¹, Nicole van der Roer¹, Dirk L. Knol^{1,4}, Heleen Beckerman^{1,4}, Maarten Boers^{1,3} & Lex M. Bouter¹

¹EMGO Institute, VU University Medical Center, Amsterdam, The Netherlands; ²Amsterdam School of Allied Health, Amsterdam, The Netherlands; ³Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands; ⁴Department of Rehabilitation Medicine, VU University Medical Center, Amsterdam, The Netherlands; ⁵EMGO Institute, VU University Medical Center, Van der Boechorststraat 7, Amsterdam, 1081 BT, The Netherlands (E-mail: hcw.devvet@vumc.nl)

Accepted in revised form 9 August 2006

Abstract

Background: Minimally important changes (MIC) in scores help interpret results from health status instruments. Various distribution-based and anchor-based approaches have been proposed to assess MIC. **Objectives:** To describe and apply a visual method, called the *anchor-based MIC distribution* method, which integrates both approaches. **Method:** Using an anchor, patients are categorized as persons with an important improvement, an important deterioration, or without important change. For these three groups the distribution of the change scores on the health status instrument are depicted in a graph. We present two cut-off points for an MIC: the ROC cut-off point and the 95% limit cut-off point. **Results:** We illustrate our *anchor-based MIC distribution* method determining the MIC for the Pain Intensity Numerical Rating Scale in patients with low back pain, using two conceivable definitions of *minimal important change* on the anchor. The graph shows the distribution of the scores of the health status instrument for the relevant categories on the anchor, and also the consequences of choosing the ROC cut-off point or the 95% limit cut-off point. **Discussion:** The *anchor-based MIC distribution* method provides a general framework, applicable to all kind of anchors. This method forces researchers to choose and justify their choice of an appropriate anchor and to define *minimal importance* on that anchor. The MIC is not an invariable characteristic of a measurement instrument, but may depend, among other things, on the perspective from which *minimal importance* is considered and the baseline values on the measurement instrument under study. A balance needs to be struck between the practicality of a single MIC value and the validity of a range of MIC values.

Key words: Health status questionnaires, Interpretation, Measurement error, Minimally important change, ROC analysis

Introduction

Health status questionnaires have become popular for measuring the effects of treatments for chronic diseases. However, changes in scores on these instruments are difficult to interpret. The statistical

significance of a change in score is partly a matter of sample size, and does not imply that the observed change is also important [1]. For clinical outcomes, such as blood pressure, clinicians have a feeling for which change is important. But an observed change in a score on a health status

questionnaire is less intuitively apparent [2]. There is a need, therefore, to define minimum changes in scores on health status questionnaires that are considered important by patients or their clinicians. A well known definition of a ‘minimally clinically important difference’ was proposed by Jaeschke et al. [3] (page 408) as ‘the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in patient management’. From a clinician’s perspective a minimally important change may be one that indicates a change in the treatment or in the prognosis of the patient [4]. Although the literature often interchanges the terms *minimally important change* and *minimally important difference*, it has been proposed that the former be used for longitudinal within-person changes in scores and the latter for cross-sectional between-person differences [5, 6]. This paper deals with minimally important change (MIC).

Crosby et al. [7] recently published an extensive overview of methods to determine MIC, distinguishing anchor-based and distribution-based approaches. In this paper, we present a visual method for determining MICs on health status questionnaires, that combines both approaches, which we call *anchor-based MIC distribution*. We first describe the method’s conceptual background, then illustrate it through an empirical example, and finally discuss its implications.

Anchor-based and distribution-based approaches

Before presenting our method, we will briefly summarize the characteristics of anchor-based and distribution-based approaches to assess MIC values as described in the elaborative review performed by Crosby et al. [7]. The anchor-based approach uses an external criterion, or *anchor* (which must substantially correlate with the health status instrument under study), to determine what patients or their clinicians consider important improvement/deterioration. Anchor-based methods assess which changes on the measurement instrument correspond with a minimal important change defined on the anchor. The advantage is that the concept of ‘minimal importance’ is

explicitly defined and incorporated in this method. All anchor-based approaches described by Crosby et al. [7] are limited in that they fail to take into account the variability of the instrument and/or the sample.

Distribution-based approaches are based on distributional characteristics of the sample, and express the observed change to some form of variation to obtain a standardized metric. Examples are effect sizes which relate observed change to the sample variability, or standardized response means which relate observed change to the variability of change. Some authors relate the observed change to the standard errors of measurement (SEM), which is a measure of the variability of the instrument [7]. The standard error of measurement quantifies the amount of error that is inherent in the instrument and/or the amount of random variation that can be expected in repeated measurements. The major disadvantage of all methods that use the distribution-based approach is that they do not, in themselves, provide a good indication of the importance of the observed change.

Therefore, Crosby et al. [7] plead for a combination of anchor-based and distribution-based methods to take advantage of both an external criterion and a measure of variability.

Combination of anchor-based and distribution-based approaches

Several authors have tried to combine the two approaches to define MICs [8–10]. Jacobson et al. [9, 10] consider patients improved once they meet both the anchor-based criterion (being closer to the point estimate of the functional mean than to the dysfunctional mean at post-test) and the distribution-based criterion (Reliable Change Index ≥ 1.96). Crosby et al. [8] determined the MIC for an obesity-specific quality of life instrument by combining the information from an anchor-based method (weight loss) and a distribution-based method (SEM corrected for regression to the mean). Without clearly stating why, they decided to consider the most conservative value as the MIC, that is, *or* the value of the anchor-based method, *or* the value of the distribution-based method.

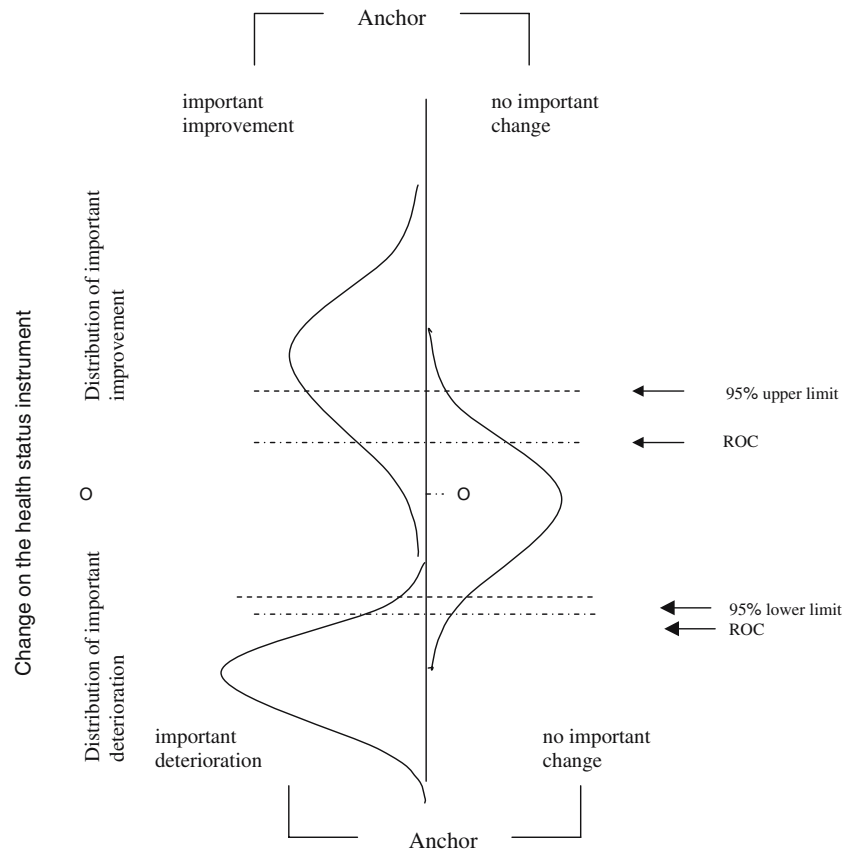


Figure 1. Distributions of the changes in scores on the health status instrument for persons who report important improvement and those who report no important change on the anchor. ROC = Receiver operating characteristic.

Presentation of the visual method: anchor-based MIC distribution

Agreeing with Crosby et al. [7], who advocate a combination of anchor-based and distribution based approaches, we not only combine the results of the two approaches, but also integrate them. We call this method *anchor-based MIC distribution*. Using an anchor, we divide a population into three groups: importantly improved, not importantly changed, and importantly deteriorated. We then plot the distribution of the change in scores on the health status instrument (Figure 1). We assess the MIC for improvement and for deterioration separately, as these can differ [7]. Next, we choose the cut-off point for an MIC. Here we will consider two cut-off points: the Receiver Operating Characteristic (ROC) cut-off point and the 95% limit cut-off point.

The ROC cut-off point is based on an ROC analysis, as applied in diagnostic studies. In this context, the health status instrument at issue is considered the diagnostic test, and the anchor functions as the gold standard [11–13]. The anchor distinguishes persons who are importantly improved or deteriorated from persons who are not importantly changed. The instrument's sensitivity is the proportion of importantly improved/deteriorated persons according to the anchor, who are correctly identified by the health status instrument as importantly improved/deteriorated. Its specificity is the proportion of 'not importantly changed' persons according to the anchor, who are correctly identified as 'not importantly changed' by the health status instrument. The ROC cut-off point is the value for which the sum of percentages of false positive and false negative classifications ($[1 - \text{sensitivity}] + [1 - \text{specificity}]$) is smallest. Note

that the assumption in this is that false positive and false negative results are equally unwanted.

The 95% limit cut-off point is based on the distribution of persons who are, according to the anchor, not importantly changed. The underlying concept is that the MIC should be detectable beyond measurement error. In other words, one might be reluctant to label persons who show no important change between the two occasions of measurement according to the anchor as importantly improved/deteriorated on the health status instrument. Using the 95% limit cut-off point, MIC for improvement is defined as the 95% upper limit of the distribution of the persons who are not importantly changed according to the anchor [mean change + 1.645 SD_{change} ¹]. Note that the 95% limit cut-off point corresponds with 95% specificity on the ROC curve.

Graphing the distribution allows one to judge how well an instrument distinguishes persons who, according to the anchor, are importantly improved or deteriorated from those not importantly changed. Moreover, the distance between the ROC cut-off point and the 95% limit cut-off point are clearly illustrated. Thus, the graph is important for seeing how the choice of a specific cut-off point influences the amount of misclassification. A flatter curve suggests a weaker correlation between anchor and health status instrument under study. Furthermore, differences in location and form of the curves of the 'improved' and 'deteriorated' persons indicate that the MICs for deterioration and improvement differ. In our theoretical example, considering the ROC cut-off points, the MIC for deterioration is larger than that for improvement, meaning that negative changes in scores must be larger than positive changes before persons think of themselves as importantly changed. Using the 95% limit cut-off point, the MIC values for improvement and deterioration are the same as long as the persons showing no important change on the anchor have a mean value of 0 on the health status instrument, and their values show a normal distribution: then both points are found at 1.96 * SD of the change scores of the *not importantly changed* group. Note that the distribution of the importantly improved/deteriorated groups have no

influence of the 95% limit cut-off point. A larger MIC for deterioration than for improvement was, for example, observed for all subscales of the Functional Assessment of Cancer Therapy instrument in cancer patients [14]. However, using an 11 point numerical rating scale to measure pain intensity, Farrar et al. [15] showed a smaller MIC for deterioration than for improvement.

Before presenting our example, we should emphasize that this *anchor-based MIC distribution* method provides a general framework, which can be applied to all kinds of anchors and definitions of *minimal importance*.

Illustration with an example

Background

We applied the *anchor-based MIC distribution* method to determine the MIC for improvement on the Pain Intensity Numerical Rating Scale (PI-NRS) in patients with low back pain (LBP) [16].

Participants

From May 2001 until December 2002 patients with non-specific LBP who were referred for physiotherapy were recruited for a randomised controlled trial, comparing an active strategy for the implementation of clinical guidelines on physiotherapy for LBP with the standard method of implementation [17]. In total, 500 patients were included.

Measures

The PI-NRS determines pain intensity on a scale from no pain (0) to very severe pain (10) [18]. The patients completed the PI-NRS at baseline and after 6, 12, 26, and 52 weeks. In this example, we use only the baseline and 12 week measurements. The patients also rated their change in health status as a global perceived effect (GPE) at 12 weeks on the following scale: (1) completely recovered; (2) much improved; (3) slightly improved; (4) no change; (5) slightly worse; (6) much worse. We used GPE as the anchor. In the primary analysis, we clustered the GPE into three categories: importantly improved (1–2), not

¹ 1.645 corresponds to 5% upper limit (one-tailed); 1.96 corresponds to 2.5% upper limit (one-tailed).

importantly changed (3–5), and importantly deteriorated (6). Only three patients fell in the latter category. This number was too small to determine the MIC for deterioration. Therefore, we excluded the three patients who were importantly deteriorated from our analyses.

Data-analysis

We compared the changes in the PI-NRS scores with the GPE categories. We considered the total sample as a cohort, ignoring the division into two treatment arms. To explore the adequateness of the anchor, we assessed the correlation (Spearman's rho) of the GPE with the changes in PI-NRS scores.

For the primary analysis, we graphed the distribution (expressed in percents) of the patients who were importantly improved (GPE categories 1–2) and those who were not importantly changed (GPE categories 3–5). To determine the ROC cut-off point for each change in PI-NRS score, we calculated the sensitivity and specificity. To construct the ROC curve, we plotted the combination of sensitivity and 1-specificity for each change in PI-NRS scores. The MIC, defined as the optimal cut-off point, is found on the ROC curve at the point closest to the upper-left corner (i.e. where the sum of the percentages of misclassified patients is lowest).

The MIC based on the 95% limit cut-off is found at the 95% upper limit of the distribution of the patients who were not importantly changed, and corresponds to the $\text{mean}_{\text{change}} + 1.645 * \text{SD}_{\text{change}}$.

To examine whether MICs differed by patient sub-group, we distinguished between patients with acute or sub-acute LBP (defined as having complaints for less than 3 months when they entered the trial) and those with chronic LBP (complaints for more than 3 months). We also performed a sub-group analysis of the baseline PI-NRS scores, defining high and low baseline values as those lying in the highest and lowest tertiles.

As a secondary analysis, we expanded the category of importantly improved to include the slightly improved patients (GPE category 3). We then graphed the distribution of the patients who were not changed (GPE category 4) and those who were slightly or more improved (GPE categories 1–3) and again determined the ROC and 95% limit cut-off points.

Results

Of the 500 participating patients 438 had complete data on the GPE and PI-NRS scores. Table 1 shows the mean changes in PI-NRS scores (with their standard deviations) for every GPE category. Spearman's rho between the changes in PI-NRS scores and the GPE categories was 0.61.

Figure 2 shows the sensitivity and specificity for various changes in PI-NRS scores. The MIC, defined as the most optimal ROC cut-off point, is at a sensitivity of 81% and a specificity of 78%, corresponding to a change in score of 2.5 points.

The 95% limit cut-off point can be calculated as $\text{mean}_{\text{change}} + 1.645 * \text{SD}_{\text{change}}$ of the not importantly changed group: $1.2 + 1.645 * 2.0 = 4.5$.

Figure 3 presents the distributions (expressed in percents) of the importantly improved and the not importantly changed patients. Both the ROC cut-off point and the 95% limit cut-off point are indicated.

Table 2, which considers patient subgroups, shows that acute and chronic patients had different MICs (for both cut-off points), and that patients with more severe pain at baseline had a greater MIC than did the patients with less severe pain.

Figure 4 presents the distributions (expressed in percents) of the importantly improved patients and the not importantly changed patients as defined in the secondary analysis. Both the ROC cut-off point and the 95% limit cut-off point are indicated. The optimal ROC cut-off point lies again at a change in score of 2.5 points. The 95% limit cut-off point can be calculated as

Table 1. The mean change scores (SD) on Pain Intensity numerical rating Scale (PI-NRS) of patients with low back pain, according to their answer on the global rating of perceived effect (anchor)

Global perceived effect	Number of patients N = 438	Change in PI-NRS Mean _{change} (SD _{change})
Completely recovered (1)	105	+5.9 (2.6)
Much improved (2)	219	+4.1 (2.4)
Slightly improved (3)	66	+1.8 (2.0)
No change (4)	28	+0.7 (2.0)
Slightly worsened (5)	17	-0.4 (1.3)
Much worsened (6)	3	-2.3 (1.5)

SD = Standard deviation.

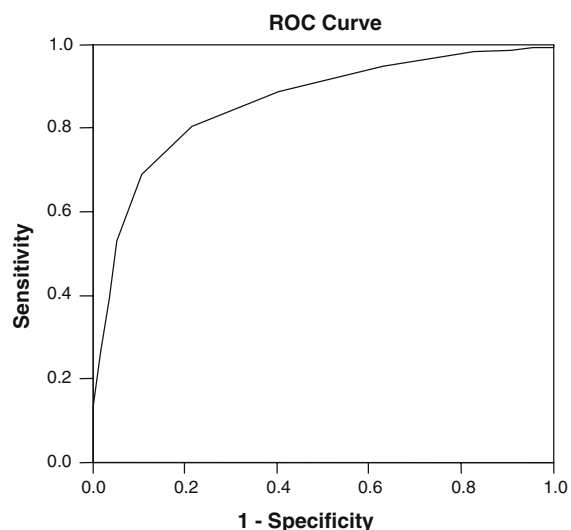


Figure 2. Receiver operating characteristic (ROC) curve for the various cut-off points for change on the Pain Intensity Numerical Rating Scale (PI-NRS), including sensitivity, specificity, and sum of percentages of misclassifications in the study of Van der Roer et al. [16].

Change on PI-NRS	True positive		True negative		Sum of % misclassified*
	N	Sens	N	Spec	
0.5	308	95%	41	37%	68
1.5	288	89%	54	59%	52
2.5	262	81%	88	78%	41
3.5	224	69%	99	89%	42
4.5	172	53%	104	94%	52

Total number of patients who were 'importantly improved': N = 324, Total number of patients with no important change: N = 111, Sens: Sensitivity, Spec: Specificity; *sum = (1-sens) + (1-spec).

$\text{mean}_{\text{change}} + 1.645 * \text{SD}_{\text{change}}$ of the not importantly changed group: $0.7 + 1.645 * 2.0 = 4.0$.

Discussion

Decisions with respect to the type of anchor

In our example we used the patient's global rating of perceived effect (GPE) as the anchor. Critics of the GPE's reliability [19] point out that it consists of only one question and that people's ability to recall their previous health status is questionable. The GPE has been shown to correlate more with

current than with previous health status [19, 20]. In our example the Spearman's rho of the GPE with the changes in PI-NRS scores was 0.61. The correlation of the GPE with the baseline and 12-week values was 0.10 and 0.80, respectively. The low correlation with baseline scores is not alarming: our study sample consisted of a homogenous group of patients who all entered the trial with severe complaints (high baseline values). During the study most patients showed a variable amount of improvement or stayed the same, leading to a more heterogeneous distribution of post-treatment values. In such a situation the correlation of the anchor with the post-treatment values will always be much higher than with the baseline values.

It is important to note that the critical remarks of using a global rating scale as an anchor do not disqualify the *anchor-based MIC distribution* method, as the method is not restricted to this specific type of anchor. Better anchors should be used if available. Cella et al. [21] present a nice example of clinical cancer outcomes as anchors, and Kolotkin et al. [22] chose change in body weight as an anchor in a study population of obese persons. Kosinski et al. [23] used five different measures for rheumatoid arthritis severity as anchor, including patient's and clinician's global assessments.

The choice of anchor is crucial in any anchor-based approach. In other words, the MIC greatly depends on the type of anchor and the anchor's definition of important change. The anchor determines whether the MIC is considered from the perspective of the patient or the clinician. As clinicians and patients do not always agree which changes are considered important the MIC from patient's perspective may differ from that from a clinician's point of view. It is fully acceptable that clinicians and patients have different perspectives on what is important: patients may base it on symptoms, and clinicians on implicit estimation of the clinical course.

Furthermore, the anchor can be very specific or quite general. A global rating scale used as an anchor, in, for example, a study on relaxation therapy for patients with angina pectoris might ask generally 'How has your health status changed since the start of the treatment?' or it might ask more specifically 'Has your anxiety deteriorated, stayed the same, or improved since the last time?'

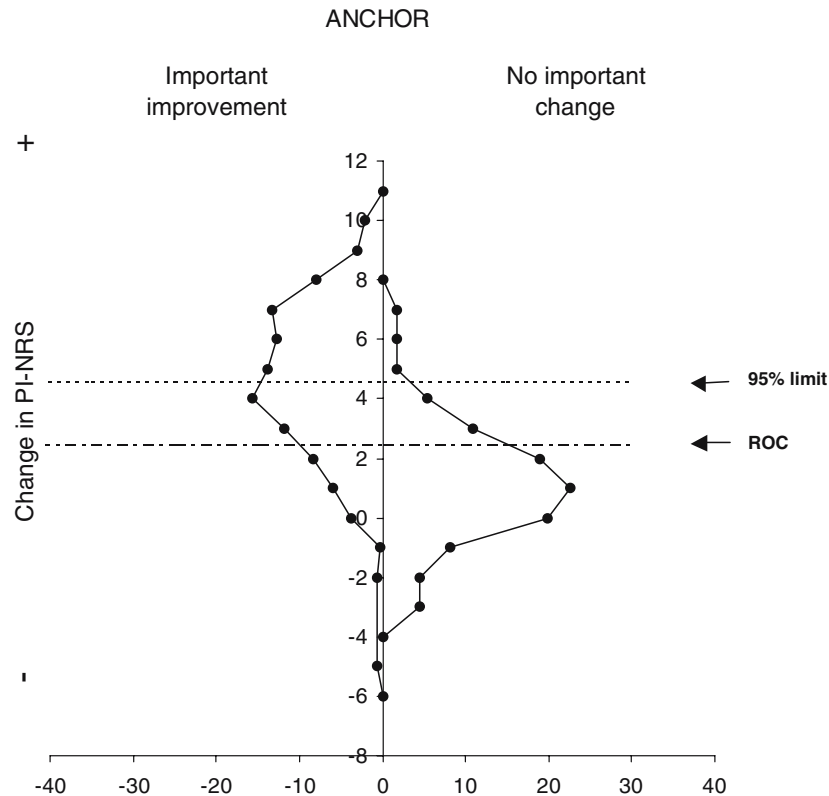


Figure 3. Distribution (expressed in percents) of changes in scores on the Pain Intensity Numerical Rating Scale (PI-NRS) for low back pain patients who report an important improvement and those who reported no important change in the study of Van der Roer et al. [16]. Slightly improved patients are considered as “not importantly changed”.

The latter question could lead to different MIC values, because anxiety is just one aspect of general health status. In general, scores on aspects of health status about which patients are less concerned must change more before they can be considered to reflect important improvement/deterioration for their health status. It has been suggested that to be an adequate anchor, it should correlate at least 0.50 with the changes in the instrument's scores [14, 24].

What is a ‘minimally important’ change?

The MIC value depends to a great degree on the anchor's definition of *minimal importance*. So, the crucial question, then, is ‘what is a minimally important improvement/deterioration?’ Some authors tend to emphasize *minimal*, while others stress *important* [25]. Remarkably little research has focused on the ‘importance’ of a change. If patients indicate to be slightly changed, it is a

Table 2. Values for minimally important change (MIC) on the Pain Intensity Numerical Rating Scale (PI-NRS) using both cut-off points in subgroups of patients with acute and chronic low back pain, and with high and low baseline values

PI-NRS ^a	Acute LBP chronic LBP		High baseline	Low baseline
	N = 300	N = 135		
ROC cut-off	2.5	3.5	3.5	1.5
95% limit cut-off	4.5	4.7	5.4	3.2

^aPI-NRS ranges from 0 points (no pain) to 10 points (very severe pain), ROC = Receiver Operating Characteristic.

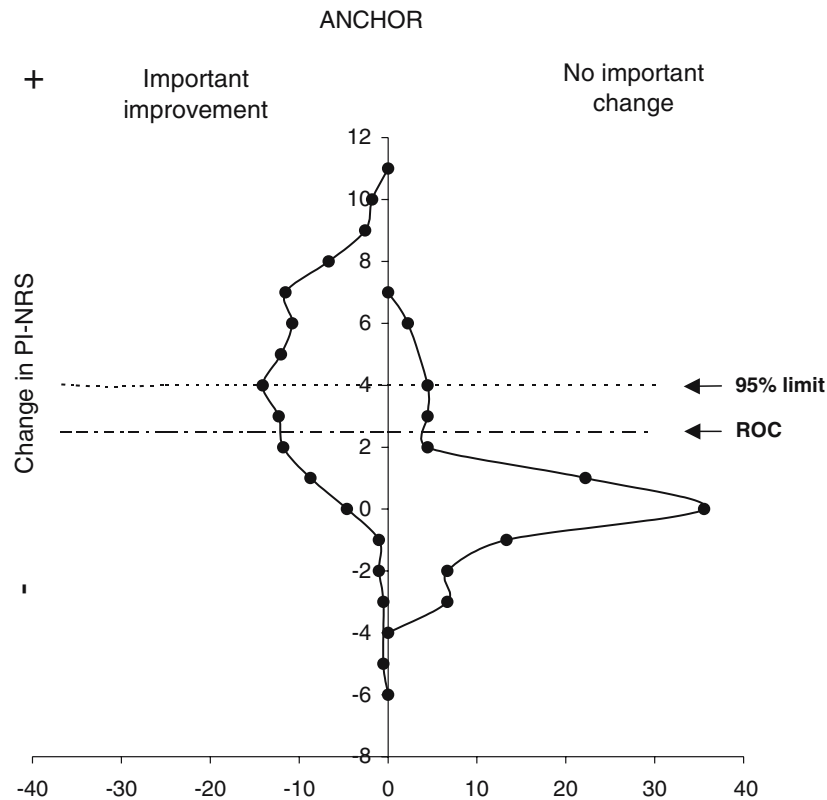


Figure 4. Distribution (expressed in percents) of changes in scores on the Pain Intensity Numerical Rating Scale (PI-NRS) for low back pain patients who report an important improvement and those who reported no important change in the study of Van der Roer et al. [16]. Slightly improved patients are considered as “importantly improved”.

minimal change but it is unknown whether this amount of change is considered important by or for these patients. A current initiative at the 8th Outcome Measures in Rheumatology (OMERACT 8) conference is aimed at exploring these issues in rheumatologic disorders (<http://www.omeract.org>).

Some authors do consider *slight improvement* as measured by the anchor to be the minimally important improvement [2, 3, 26]. We [16, 27, 28] and others [15, 29–31] set the bar for minimally important improvement at *much improved*. We had several reasons for this choice in our primary analysis. In our opinion, it better reflects the concept of important improvement, and we expect that some patients, wanting to please their doctor or researchers, easily say that they are slightly improved.

In our secondary analysis we did lower the bar for minimally important improvement to include

those persons who indicated on the anchor that they had *slightly improved*. In that analysis, the MIC using the ROC cut-off was again 2.5, but the MIC value using the 95% limit cut-offpoint was somewhat smaller, and the overlap between the two curves was substantially larger. This overlap, however, says nothing about the most adequate definition of minimally important improvement, which, in its very nature, is arbitrary.

Which cut-off point is preferred?

A challenging question is: Should the ROC cut-off point or the 95% limit cut-off point be used as the MIC? With the ROC cut-off point, false positive and false negative classifications are equally weighted. If there is no *a priori* reason to dislike false positives more than false negatives, the ROC cut-off point is a good choice. However, if one objects to classify patients as improved whose

changes in scores fall within the measurement error of the not importantly changed patients, one might prefer the 95% limit cut-off point. Alternative cut-off points are also defensible, as long as a justification is given.

We recommend graphs of the *anchor-based MIC distribution* to visualize the consequences of both ROC and 95% limit cut-off points. The ROC cut-off point usually results in a smaller MIC value than the 95% limit cut-off point, meaning that less change is needed before it is considered important. Note that in Figure 1, in the assessment of the MIC for deterioration, the ROC curve cut-off is larger (i.e. larger distance from zero) than the 95% cut-off level. This can only be reached if the curves hardly overlap, in other words, the optimal cut-off point on the ROC curve has a specificity of more than 95%.

MIC is not an invariable characteristic

Some authors have advocated one uniform measure for MIC, such as 0.5 points on a 7-point response scale [2] or one SEM [32, 33]. Other studies, using an anchor-based method, however, have shown that an MIC is not an invariable characteristic. It depends on baseline values – with higher baseline values (more severe disorders) needing greater changes to be labeled important [8, 31, 34, 35] – and even on characteristics such as age and sex [36]. What is considered to be an MIC depends, among other things, on the anchor, on the severity of the disease, and on the intervention.

To investigate whether sub-groups of patients require different MICs, we calculated the MICs for subgroups of (sub)-acute and chronic patients, and for patients with high and low baseline values. An accommodation for MICs' dependency on baseline values is to express the MIC as a percentage of baseline values. Farrar et al. [15] showed that MICs for a pain intensity rating scale were more uniform when expressed as percentage of baseline values than as absolute change. This solution, unfortunately, does not apply to other characteristics that may affect MICs.

How to deal with different values for MIC

Once it is acknowledged that an MIC cannot be expressed as a single value, it follows that it should

be expressed as a range that includes all reasonable values [23, 37, 38]. Ranges, however, require that people know when to use the larger values and when the smaller. People will tend to choose the smallest MIC – they want, after all, see improvement – but the smallest value may not be the most adequate in their situation. In case of high baseline values, for example, higher MICs apply. It is the challenge to balance the clinical practicality of an easily applied single value against the validity of a harder-to-determine value within a range. We support the view of Sloan et al. [25] that, for MICs to be accepted and used in clinical practice, a single value should be set, but with a small range around it to accommodate some variation. As in the end the MIC should be viewed as a tool to improve interpretation of study (or measurement) results, strongly based on perceptions of those involved, there is a good case to use a mix of evidence-based and consensus processes to come to reasonable and parsimonious choices on MIC values. The OMERACT initiative has been highly successful in organizing such processes in the field of rheumatology (see: <http://www.omeract.org>). These initially set MICs can always be moved if further research so demands.

The MIC, though important, is only one of the values that enhance our interpretation of the scores on health status instruments. Comparing scores from different patients groups [39] and relating scores to other, better understood, clinical parameters [23] also enhance the interpretation of these instruments. Our Table 1 is informative in that respect.

Relation of the anchor-based MIC distribution method with other methods for assessing MIC

Authors such as Juniper et al. [2] and Farrar et al. [15] have defined the MIC as the mean change in scores of patients categorized by the anchor as having experienced minimally important improvement/deterioration. As can be seen in Table 1, when minimally important improvement was set at *much* improved, the patients that fell within the category had a mean score of 4.1. When the bar was lowered to *slightly* improved the mean score of persons in that category was 1.8. Note that this method does not take into account the standard deviation of these changes in scores, and

only the category of minimally important improvement is used.

Including the categories of improvement beyond minimally important would falsely increase the MIC, because patients who are considered completely recovered are more likely to have very high changes in scores. However, for the ROC analysis, considering only the category of minimally important improvement underestimates the number of false negative classifications, because the categories that indicate more than minimally important improvement may include persons who score lower than the optimal ROC cut-off point. One certainly wants to define these as false negatives. Therefore we have sub-divided our total sample (except for the three deteriorated persons) into importantly improved and not importantly changed persons to determine the minimal important change.

With respect to the role of the distribution, also the ROC analysis ignores standard deviations or other distribution parameters. The ROC cut-off point is based on the minimum percentages of misclassifications on the health status instrument with the anchor as gold standard.

The standard deviation of changes on the health status instrument first becomes important if the 95% limit cut-off point is used. Note that in that case, one only considers the distribution of the persons who have not experienced minimally important change.

Many authors proposed distribution-based approaches to assess MIC, most of which express the observed change in a standardized metric. The SEM, an often-used distribution-based measure, links the reliability of the health status instrument to the standard deviation of the population [7]. The major disadvantage of all distribution-based methods is that they reveal minimally *detectable* change rather than minimally *important* change; in themselves, they cannot provide a good indication of the importance of the observed change. Although it may appear, at first glance, to make sense to define an MIC on what is detectable, this leads to the faulty reasoning that what is detectable is important, and conversely, that what is undetectable cannot be important. The latter reasoning has the unfortunate effect of making it impossible to ever conclude that an instrument is unsuitable for detecting MICs.

Statistical significant changes on group level, on individual level, and MIC

It is widely acknowledged that statistically significant differences on group level are largely dependent of sample sizes and have little relation to MICs for individual patients. A variety of approaches to determine the statistical significance of individual change have been proposed [40]. Our 95% limit cut-off point incorporates the concept of statistical significance of individual change, representing a change that is statistically significant different from persons who do not importantly change. The ROC cut-off point is more liberal in this respect, and may result in MIC values which are not statistically different from the mean value of the patients that do not experience an important change.

To use the MIC values on group level, for example to interpret the results of clinical trials, one should determine the proportion of patients who show changes larger than the MIC in each treatment group and compare these proportions [41, 42].

Conclusion

The *anchor-based MIC distribution* method truly integrates the anchor-based and distribution-based approaches, thus taking advantage of an anchor with measures of precision to establish cut-off points that are interpretable and based on a desired confidence level.

The *anchor-based MIC distribution* approach provides a general framework, applicable to all kind of anchors. The definition of *minimal important change* is not an inherent characteristic of the method. However, it forces researchers to choose and justify their choice of an appropriate anchor and to define *minimal importance* on that anchor.

The method's graphical presentation shows the adequateness of the anchor and the consequences of choosing a specific MIC.

References

1. Wright JG. The minimal important difference: Who's to say what is important. *J Clin Epidemiol* 1996; 49: 1221–1222.

2. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol* 1994; 47: 81–87.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–415.
4. Walraven C van, Mahon JL, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: Implications for sample-size calculation. *J Clin Epidemiol* 1999; 52: 717–723.
5. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *J Rheumatol* 2001; 28: 400–405.
6. Vet HC de, Beckerman H, Terwee CB, Terluin B, Bouter LM. Definition of clinical differences. *J Rheumatol* 2006; 33: 434.
7. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56: 395–407.
8. Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. *J Clin Epidemiol* 2004; 57: 1153–1160.
9. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59: 12–19.
10. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *J Consult Clin Psychol* 1999; 67: 300–307.
11. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986; 39: 897–906.
12. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991; 12: 142S–158S.
13. Stratford PW, Binkley FM, Riddle DL. Health status measures: Strategies and analytic methods for assessing change scores. *Phys Ther* 1996; 76: 1109–1123.
14. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: Differences between improvement and worsening. *Qual Life Res* 2002; 11: 207–221.
15. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001; 94: 149–158.
16. Roer N van der, Ostelo RW, Bekkering GE, Tulder MW van, Vet HC de. Minimal clinically important change for different outcome measures in patients with nonspecific low back pain. *Spine* 2006; 31: 578–582.
17. Bekkering GE, Tulder MW van, Hendriks EJ, Koopmanschap MA, Knol DL, Bouter LM, Oostendorp RA. Implementation of clinical guidelines on physical therapy for patients with low back pain: Randomized trial comparing patient outcomes after a standard and active implementation strategy. *Phys Ther* 2005; 85: 544–555.
18. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: A comparison of six methods. *Pain* 1986; 27: 117–126.
19. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol* 2002; 55: 900–908.
20. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: The lesson of Cronbach. *J Clin Epidemiol* 1997; 50: 869–879.
21. Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, Wolf MK, Johnson DH. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 2002; 55: 285–295.
22. Kolotkin RL, Crosby RD, Williams GR. Integrating anchor-based and distribution-based methods to determine clinically meaningful change in obesity-specific quality of life. *Qual Life Res* 2002; 11: 670.
23. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum* 2000; 43: 1478–1487.
24. Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new drugs. *Pharmacoeconomics* 1997; 12: 621–626.
25. Sloan JA, Cella D, Hays RD. Clinical significance of patient-reported questionnaire data: Another step toward consensus. *J Clin Epidemiol* 2005; 58: 1217–1219.
26. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; 37: 469–478.
27. Beurskens AJ, Vet HC de, Koke AJ. Responsiveness of functional status in low back pain: A comparison of different instruments. *Pain* 1996; 65: 71–76.
28. Ostelo RW, Vet HC de, Knol DL, Brandt PA van den. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol* 2004; 57: 268–276.
29. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004; 8: 283–291.
30. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther* 1999; 79: 371–383.
31. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 1. *Phys Ther* 1998; 78: 1186–1196.
32. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying

- meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861–873.
33. Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res* 2002; 11: 1–7.
 34. Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; 12: 12–20.
 35. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2. *Phys Ther* 1998; 78: 1197–1207.
 36. Santanello NC, Zhang J, Seidenberg B, Reiss TF, Barber BL. What are minimal important changes for asthma measures in a clinical trial?. *Eur Respir J* 1999; 14: 23–27.
 37. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD: J Chronic Obstructive Pulmonary Dis* 2005; 2: 63–67.
 38. Ostelo RW, Vet HC de. Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol* 2005; 19: 593–607.
 39. Wolfe F, Michaud K, Strand V. Expanding the definition of clinical differences: from minimally clinically important differences to really important differences. Analyses in 8931 patients with rheumatoid arthritis. *J Rheumatol* 2005; 32: 583–589.
 40. Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui K. Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Eval Health Professions* 2005; 28: 160–171.
 41. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77: 371–383.
 42. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *Br Med J* 1998; 316: 690–693.

Address for correspondence: Henrica C.W. de Vet, EMGO Institute, VU University Medical Center, Amsterdam, The Netherlands
Phone: +31-20-444-8176; Fax: +31-20-444-6775;
E-mail: hcw.dev@vumc.nl