Published in final edited form as:

J Clin Epidemiol. 2014 June ; 67(6): 689–696. doi:10.1016/j.jclinepi.2013.10.025.

DEPENDENCE OF THE MINIMAL CLINICALLY IMPORTANT IMPROVEMENT ON THE BASELINE VALUE IS A CONSEQUENCE OF FLOOR AND CEILING EFFECTS AND NOT DIFFERENT EXPECTATIONS BY PATIENTS

Michael M. Ward, MD, MPH,

Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Lori C. Guthrie, RN, BSN, and

Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Maria Alba, MD

Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Abstract

Objective—Estimates of minimal clinically important improvements (MCII) are larger among patients with higher values at baseline, suggesting these patients require larger changes to appreciate improvements. We examined if baseline dependency of MCII was associated with specific patients across three measures, or was due to floor and ceiling effects.

Study design and setting—We prospectively examined 250 outpatients with active rheumatoid arthritis (RA). We used an anchor-based approach to estimate MCII for three measures of RA activity (patient global assessment, swollen joint count, walking time). We examined if the same patients constituted the baseline subgroups with high MCII across measures.

Results—MCII were greater for those with higher baseline values of all three measures. At the ceiling, there was little opportunity to improve and judgments were unrelated to measured changes. At mid-range, improvements were balanced by worsenings, including some judged as improvements. At the floor, improvements were not similarly balanced. Patients in subgroups with high MCII for patient global assessment were not also predominantly in subgroups with high MCII for the swollen joint count or walking time, and vice versa.

Address for correspondence: Michael M. Ward, MD, NIAMS/NIH, Building 10 CRC, Room 4-1339, 10 Center Drive, Bethesda, MD 20892. Telephone (301) 496-7263; Facsimile (301) 480-2714; wardm1@mail.nih.gov.

Financial Disclosures: M. M. Ward none; L. C. Guthrie none; M. Alba none

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusion—Variation in MCII by baseline values is due to floor and ceiling effects rather than expectations of particular patients.

Keywords

Minimal clinically important difference; responsiveness; ceiling effect; floor effect; outcome measures

INTRODUCTION

Patient-reported outcomes have become recognized as central components in the assessment of health, and are now routinely included as endpoints in clinical trials and observational studies. While the comparison of responses between treatment groups provides an estimate of the effects of treatment, this comparison does not provide information on whether the improvement was substantial or trivial. Full interpretation requires knowing what degree of change in a measure represents an important or clinically-meaningful change, and whether a higher proportion of patients in one group met this threshold.[1] In addition to facilitating the interpretation of trial results, the minimal clinically important improvement (MCII) of a study's primary outcome is important in study design as a guide to sample size estimation. Although the MCII has most often been assessed for patient-reported outcomes, similar issues pertain to measures that are not patient-reported.

Of several approaches used to estimate the MCII, anchor-based methods are the most direct, and frequently use the patient's explicit judgment of improvement as an external standard. [2] Most often, investigators intend to determine a single MCII for a given measure. However, the MCII may vary with the analytic approach or the nature of patients assessed. [3,4] Several studies have examined potential sources of variation in the MCII, including for example whether the MCII was similar for men and women, as an indication of whether group-specific MCIIs were needed.[5] A notable observation has been that when patients are stratified by their value on the measure at study baseline, estimates of the MCII are substantially larger for subgroups of patients with high baseline values (or values indicating more severe disease) than for subgroups with midrange or low values (or values indicating less severe disease). Dependence of the MCII on the baseline value was observed in each of the 27 studies we identified that examined the baseline value as a source of variation in the MCII.[5–31] This dependence was irrespective of the nature of the outcome, which ranged from pain scales and functional indices to urinary symptom scales, and irrespective of the format of the measure, suggesting that it may be axiomatic.

This association has commonly been interpreted to indicate that patients with more severe symptoms require a larger improvement to appreciate that they are better than those with less severe symptoms. Although this interpretation is logical, the universality of this association across studies, conditions, and measures suggests that the dependency of the MCII on the baseline value may be a consequence of the measurement process, rather than a truism of how patients perceive health changes. Most measures are bounded, and improvements, by definition, are unidirectional. Floor and ceiling effects have been invoked as possibly contributing to this observation, but this possibility has not been explored in

detail.[8,19,32] In this study, we examined whether floor and ceiling effects might account for the baseline dependency of the MCII in a study of patients with active rheumatoid arthritis (RA). We examined three different measures of RA activity: the patient global assessment, a widely-used patient-reported measure of overall arthritis activity; the swollen joint count, a physician-derived measure; and walking time, a performance measure. In addition to testing if the MCII varied with the baseline value of each measure, we examined if the same subset of patients was identified as having a high MCII for each RA activity measure. We hypothesized that if baseline dependency of the MCII was due to the “requirements” or judgments of a particular subgroup of patients, the same subgroup should be identified by each measure of RA activity. In contrast, if each measure identified different sets of patients as having a high MCII, the baseline dependency of the MCII would relate to the measure rather than to the requirements or expectations of particular patients.

METHODS

Subjects

We enrolled subjects with RA in a prospective longitudinal study of changes in RA activity with treatment. The goal of the study was to determine the sensitivity to change and MCII for commonly-used measures of RA activity. Eligible subjects were age 18 years or older, met classification criteria for RA [33], and had active RA with at least 6 tender joints and in the judgment of their rheumatologist required an escalation in treatment with either disease-modifying medications or systemic corticosteroids. The choice of treatment was left to the treating rheumatologist and not dictated by the study. Subjects were recruited from the outpatient clinics of the investigators.

Study procedures

Subjects were invited to participate at the time treatment was escalated or within 3 days of escalation. After obtaining written informed consent, we performed a clinical assessment of the subject, which included a complete joint count for tenderness (68 joints) and swelling (66 joints). We also collected patient-reported outcomes, including a patient global assessment by visual analog scale (0 = very well; 100 = very poor). To assess the importance they ascribed to different aspects of RA, we asked subjects to identify the three aspects (from a list of 10, including joint swelling and functional difficulty) that were most important to them and that they would most want to resolve. Additionally, we tested walking ability by timing with a stopwatch how fast the subject could walk 50 feet in a hallway. Low (more normal) scores on each measure were considered the ceiling.

Assessments were repeated at a second study visit, which occurred one month after the initial visit if the treatment change was escalation in systemic corticosteroids, and four months after the initial visit if the treatment change was escalation in disease-modifying medications. The difference in timing of the second assessment was needed because responses to systemic corticosteroids occur faster than those to disease modifying medications. The same study investigator performed both assessments of individual subjects.

At the second visit, subjects were asked to complete a questionnaire that included questions on whether they considered that overall their arthritis had improved, worsened, or stayed the same. If either improved or worsened, they were asked to rate the importance of the change on a 7-point scale (almost none, hardly important at all; a little important; somewhat important; moderately important; a good deal important; very important; extremely important).[34] Similar anchor questions were asked for improvement in joint swelling and ability to do things.

Statistical analysis

To compute the MCII, we constructed receiver operating characteristic (ROC) curves for each measure, distinguishing subjects who reported any improvement in the anchor question from those who reported no change or any worsening.[35] We chose as the MCII the change score associated with a specificity for improvement of 0.80 or more, following previous work.[25] Using this method to derive the MCII facilitated comparisons among measures and subgroups because the sensitivity of all MCII was fixed. To examine how the MCII varied with the baseline value of the patient global assessment, we computed the MCII in 10 subgroups defined by 10 point increments in the baseline value, ranging from those with a baseline value of 0 – 9 (very well) to those with a baseline value of 90 – 100 (very poor). We computed the MCII for the swollen joint count in 7 subgroups, ranging from those with a baseline count of 5 or fewer swollen joints to those with a baseline count of 31 or more swollen joints. We computed the MCII for walking time in 5 subgroups, defined by quintiles of the baseline, ranging from 8 seconds or less in the best-performing quintile to 15 seconds or more in the worst-performing quintile.

To examine the association between the MCII and ceiling and floor effects, we plotted the change in each measure by baseline subgroup, and noted the occurrence of counterintuitive responses to the anchor question in each subgroup. A counterintuitive-positive response was considered one in which the measure worsened between assessments but the subject judged himself as improved. A counterintuitive-negative response was considered one in which the measure improved between assessments but the subject judged himself to have worsened or stayed the same.

Because ceiling effects can limit the sensitivity to change of measures, we also examined the relationship between the MCII and sensitivity to change. We assessed sensitivity to change using standardized response means (mean change/standard deviation of the change), and computed these for each baseline group.[36]

To determine if a high MCII identified the same subset of subjects by each measure, we examined the concordance of baseline values among the three RA activity measures. We also examined whether patients in the highest MCII subgroups for the patient global assessment were also concentrated in the highest MCII subgroups for the swollen joint count and walking time, as a direct assessment of the overlap of these groups.

There were no missing data for the patient global assessment or swollen joint count. Walking time was missing at either the baseline or follow-up assessment in 18 subjects, and these were excluded from analyses that used walking time.

We used SAS programs, version 9.3 (SAS Institute, Cary, NC) for all analyses.

RESULTS

Subjects

We studied 250 subjects (78% women; mean ($\pm$ standard deviation) age 51.1 ± 13.7 years; median duration of RA 6.4 years). At baseline, the mean patient global assessment was 55.6 ± 25.2 , with a range from 3 to 100. The median swollen joint count was 15, with a range from 2 to 46. The mean walking time was 12.6 ± 5.9 seconds, with a range from 6 to 51 seconds. Therefore, although subjects fulfilled the inclusion criteria for active arthritis, there was a wide range of ratings on these three measures of RA activity. Fifty-six subjects (22.4%) were treated with systemic corticosteroids and had their second visit at 1 month, while the remainder ($N = 194$) had escalation of disease-modifying medications and their second visit at 4 months. At the second visit, 167 subjects (66.8%) reported that their overall arthritis status had improved, 158 subjects (63.2%) reported that their joint swelling had improved, and 151 subjects (60.4%) reported that their functional ability had improved. These improvements were rated as at least moderately important by 92%, 93.6%, and 95.4% of subjects on the three measures, respectively. Eighty-two percent of patients assessed at 1 month reported improvement in overall arthritis status, compared to 63% assessed at 4 months, but the proportion of patients assessed at 1 month was not associated with the baseline subgroup.

Associations with Patient Global Assessment

Among all subjects, a decrease of 19 points or more on the 0 – 100 patient global assessment scale had a specificity of 0.80 for a judgment of improvement in overall arthritis status, based on a ROC curve with an area of 0.74 (95% confidence interval (CI) 0.68, 0.81). The correlation between measured changes in the patient global assessment and patients' ratings of subjective change was 0.36.

When stratified by the patient global assessment at study baseline, the MCII was -42 in the subgroup with a baseline value of 90 – 100, and was lower among subgroups with lower baseline values (Table 1). For example, the MCII was -29 for the subgroup with a baseline value of 60 – 69, and -7 for the subgroup with a baseline value of 30 – 39. These results are consistent with those of previous studies showing an association between the baseline value and MCII. Subgroups with a baseline value of either 0 – 9 or 10 – 19 had the improbable result that the estimated MCII was a positive value (indicating worsening, rather than improvement).

The pattern of changes and corresponding judgments of improvement by baseline subgroup reveals several important findings (Figure 1). First, the maximum possible improvement was dictated by the baseline value, and changes are distributed widely among all possible improvements. Subjects with a higher baseline score have an opportunity to register a greater potential improvement than those with a lower baseline score. Second, at the floor (80 – 100), where there was very little opportunity for scores to worsen, counterintuitive-negative responses occurred, but counterintuitive-positive responses were not observed or

were very rare. Third, counterintuitive-negative responses that occurred among subjects at the floor tended to be associated with large improvements in the measure, because large improvements were common and worsening was restricted. This had the effect of inflating the MCII estimate at the floor. For example, if the three most extreme counterintuitive-negatives in the 90–100 subgroup had instead been judged as improvements, the MCII in this subgroup would have been –19 rather than –42. Fourth, in the mid-range baseline subgroups, although many subjects improved, worsening was also observed. While some counterintuitive-negative responses were observed in these subgroups, more of the negative judgments were associated with worsened scores, as would be expected. True negative responses were not observed to occur in the subgroups at the floor, because worsening was restricted. Fifth, counterintuitive-positive responses also occurred in the mid-range baseline subgroups, which were not observed in the subgroups at the floor. The more limited range for possible improvement, the lower prevalence of counterintuitive-negatives, and the higher prevalence of both worsenings and counterintuitive-positives combined to have the effect of decreasing the MCII estimates in the mid-range baseline subgroups relative to those at the floor. Sixth, in subgroups at the ceiling (0 – 20), the observed improvements were small and balanced by worsenings, and associated judgments appeared to be random. The result was the finding that increases (worsening) in the patient global assessment were associated with judgments of improvement.

The estimate of the MCII by subgroup paralleled the standardized response means (Table 1). At the ceiling, the patient global assessment was poorly sensitive to change, which underlies the poor agreement between measured changes and patient judgments of improvement in these subgroups.

Associations with Swollen Joint Count

Among all subjects, the MCII for the swollen joint count was a decrease in 7 swollen joints, based on a ROC curve with an area of 0.72 (95% CI 0.65, 0.79). The correlation between measured changes in the swollen joint count and subjective ratings of change was 0.38. Results of the analysis of subgroups stratified by the swollen joint count at baseline were closely similar to those for the patient global assessment. The MCII estimate was a decrease in 22 swollen joints for the subgroup with more than 30 swollen joints at baseline, and was progressively lower for subgroups with lower values at baseline (Table 1). For the subgroup with five or fewer swollen joints at baseline, the improbable result of an increase in two swollen joints as the MCII was observed. The SRM decreased with the baseline joint count, and was not sensitive to change for those with five or fewer swollen joints. The pattern of counterintuitive responses at the floor, and counterintuitive-positive responses and worsenings in the mid-range baseline groups, was similar to that seen for the patient global assessment (Figure 2). At the ceiling, there was little apparent association between measured changes in the number of swollen joints and judgments of improvement. The subgroups with high MCII were no more likely than those with low MCII to identify joint swelling as one of the three manifestations of RA they would most want to resolve, suggesting that differences in valuation of individual aspects of RA likely did not have a role in the MCII differences.

Associations with Walking Time

The MCII estimate for walking time among all subjects was a decrease of 2.8 seconds, based on a ROC curve with an area of 0.64 (95% CI 0.56, 0.71). The correlation between measured changes in walking time and subjective ratings of change in ability to do things was 0.27. Associations were again present between the baseline walking time and both the MCII estimate and SRM, with larger values among subjects with longer walking times at baseline (Table 1). For the most mobile subgroup, the SRM indicated poor sensitivity to change and the MCII estimate was 0 seconds. Although walking time is not strictly bounded at the floor, few subjects with the longest walking time at baseline worsened with treatment (Figure 3). Subjects in this subgroup had greater opportunity to demonstrate improvement, and similar to the other measures, counterintuitive-negative responses were observed in some patients who had large improvements in walking time. In the mid-range baseline subgroups, opportunity to demonstrate improvement was more limited and accompanied by counterintuitive-negative, and less often, counterintuitive- positive, responses. There was also no association between the walking time subgroups and the proportion that rated physical limitations as one of the three aspects of RA most wanted resolved.

Concordance of Subjects at the Floor on Different Measures

We examined if subjects with a high MCII for the patient global assessment were the same subjects who constituted the subgroups that had a high MCII for the swollen joint count and walking time, by plotting the joint distribution of baseline values by these measures. Those with a patient global assessment at baseline of 90 – 100 and of 80 – 90 had widely varying baseline values for the swollen joint count and for walking time (Figure 4). This indicates that while as a group they had a high apparent MCII for the patient global assessment, they were also part of groups that did not have high MCIIs for either the swollen joint count or walking time. Similar wide distributions across patient global assessments and walking times were observed for those with high swollen joint counts at baseline, and for patient global assessments and swollen joint counts among those with high walking times at baseline (Supplemental Figure 1 and 2).

Lastly, we examined if patients in the subgroups with the highest MCII for the patient global assessment were also concentrated among the subgroups with the highest MCII for the swollen joint count and walking time (Figure 5A and 5B). Patients with a global assessment of 80 to 100 at baseline were equally likely to be represented among subgroups with low and higher MCII for the swollen joint count and walking time. In addition, patients in this subgroup did not have more extreme improvements in joint swelling or walking time than those with lower baseline values of the patient global assessment, relative to other subjects in the same baseline subgroup of swollen joint count and walking time. These results indicate that subjects in the most severely affected category on one measure do not similarly require large improvements on other measures to judge themselves as improved.

DISCUSSION

Consistent with previous literature on other measures and conditions, estimates of the MCII for the three measures of RA activity we studied varied with the baseline value of the

measure.[5–31] Subgroups with more severe values on each measure at study entry had a higher MCII than those with less severe values at study entry. This finding was present not only for the patient-reported measure but also for a physician-reported measure and a performance measure, despite differences in their formats and gradation. However, in comparing responses across the three measures, we also found that subjects who comprised the subgroups with a high MCII for the patient global assessment did not also comprise the subgroups with high MCII for the swollen joint count or walking time, and vice versa. This pattern of associations would not be expected if particular patients with more severe RA “required” a greater improvement or judged changes differently than those with less severe RA. If this was the case, we would anticipate that the same patients would have similarly strict expectations for improvement across the three measures. The poor concordance among patients on the three measures at baseline and in the distribution of their responses suggests that characteristics of particular patients are not responsible for the baseline dependency of the MCII. Rather, the MCII associations with the baseline value segregated with the measure rather than with the patient.

Our results suggest that the baseline dependency of the MCII is due to floor and ceiling effects, rather than the nature of particular patients. To appreciate the consequences of floor and ceiling effects, it is helpful to first examine the determinants of the MCII in the mid-range baseline groups, which are less affected by floor or ceiling effects. In these groups, improvements are not only bounded by a maximum possible change, but improvements in some patients are balanced by worsening in other patients despite treatment. Overlaid on this variation in measured changes is variability in patient judgments of improvement. Although many patients with measured improvement judged themselves as improved, others with similar degrees of measured improvement judged themselves as not improved. Similarly, some patients with measured worsening judged themselves as improved. These factors together produce an estimate of the MCII that incorporates any error in measurement or in patient judgment or memory of their pre-treatment status. These factors are altered at the floor. Not only is there greater opportunity to register improvement at the floor, but misjudgments of no improvement are more typically associated with larger measured improvements, which inflates the MCII. Additionally, because the potential for worsening is constrained at the floor, improvements are not balanced by the results of patients who have measured worsening but who judge themselves to be improved. Therefore, the MCII estimate at the floor will be larger than the MCII estimate at the mid-range. At the ceiling, the maximum possible improvement is limited. When combined with variations in patients’ judgments, MCII estimates at the ceiling are lower than those at the mid-range.

The MCII estimates at the ceiling for each of the three measures were not only lower than those at the mid-range, but measured worsening or no change was designated as the MCII. The improbability of similar results has been raised previously.[7] This result is a consequence of both conceptual and operational limitations induced by the ceiling. Conceptually, the MCII is not relevant for patients who are already doing well. These patients have passed a threshold where further improvement is difficult to perceive reliably and judgments regarding improvement become difficult. Operationally, to compute a valid MCII, a measure must have sufficient capability to register changes should they occur. Sensitivity to change is thus a prerequisite for determining the MCII. Limited sensitivity to

change contributes to the low estimates of the MCII at the ceiling. Estimation of the MCII should be reserved for those meeting some minimum level of clinical severity, and for those with sufficient capability to register a change.

Although we targeted patients with active RA, this study was facilitated by the wide range of baseline measures among patients. Testing several measures in the same group of patients allowed us to isolate the effects of the measures from the characteristics of the patients. Although MCII might be expected to differ among subgroups if some patients valued certain aspects of RA differently than others, we did not find that subgroups with high MCII for joint swelling and walking time valued improvement in these measures more often than those in subgroups with lower MCII. By testing different types of measures, we could also show that baseline dependency was not limited to patient-reported measures. The ROC curve area for walking time was admittedly not large, but judgments of improvement were nonetheless significantly associated with measured improvements in walking time. We focused on the ROC curve approach to MCII estimation, as it is currently the most commonly used method. However, similar floor and ceiling effects would be expected to contribute to MCII estimates based on mean changes, without considering the effects of miscategorized judgments. We used a judgment of any improvement as the threshold to calculate the ROC curve. Although an argument could be made for using a higher threshold to define important improvement, the large majority of those who rated themselves as improved reported that the improvement was at least moderately important. Our study is limited in that we used only one anchor to derive each MCII estimate, raising the possibility that results may have been different if other anchors had been used. While the findings were consistent for the three measures, future studies might include additional measures, and examine these relationships in other diseases. Qualitative methods could also be used to investigate influences on patient's expectations for improvement.

Based on these results, and in the context provided by prior literature, we recommend that dependence of the MCII on the baseline value be accepted as a feature of the measurement process and not as a distinction of individual groups of patients. Although the MCII is larger among subgroups with more severe disease at baseline, the conclusion that patients with more severe disease at study baseline require greater improvement than those with less severe disease in order to judge themselves as improved is a misinterpretation. Because baseline values do not identify patients with unique requirements or expectations, there is no need to specify different MCII for different subgroups of patients. Rather, a single MCII can be estimated for all patients who satisfy a target level of clinical activity or severity for the condition being studied. Patients who do not meet the minimal target level of clinical severity should be excluded from calculation of the MCII to avoid attenuation of the estimate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, and U.S. Public Health Service grant AR45177.

References

1. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002; 77:371–83. [PubMed: 11936935]
2. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003; 56:395–407. [PubMed: 12812812]
3. Testa MA. Interpretation of quality-of-life outcomes: Issues that affect magnitude and meaning. *Med Care.* 2000; 38 (Suppl II):66–74.
4. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for examining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008; 61:102–9. [PubMed: 18177782]
5. 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; 94:149–58. [PubMed: 11690728]
6. Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker-Corkery E, et al. Benign prostatic hyperplasia specific health status measures in clinical research: How much change in the American Urological Association Symptom Index and the Benign Prostatic Hyperplasia Impact Index is perceptible to patients? *J Urol.* 1995; 154:1770–4. [PubMed: 7563343]
7. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. *Phys Ther.* 1996; 76:359–65. [PubMed: 8606899]
8. Baker DW, Hays RD, Brook RH. Understanding changes in health status: Is the floor phenomenon merely the last step of the staircase? *Med Care.* 1997; 35:1–15. [PubMed: 8998199]
9. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 1. *Phys Ther.* 1998; 78:1186–96. [PubMed: 9806623]
10. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: Part 2. *Phys Ther.* 1998; 78:1997–207.
11. Bird SB, Dickson EW. Clinically significant changes in pain along the visual analog scale. *Ann Emerg Med.* 2001; 38:639–43. [PubMed: 11719742]
12. Jensen MP, Chen C, Brugger AM. Interpretation of visual analog scale ratings and change scores: A reanalysis of two clinical trials of postoperative pain. *J Pain.* 2003; 4:407–14. [PubMed: 14622683]
13. Hägg 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. [PubMed: 12592542]
14. Cepeda MS, Africano JM, Polo R, Alcalá R, Carr DB. What decline in pain intensity is meaningful to patients with acute pain? *Pain.* 2003; 105:151–7. [PubMed: 14499431]
15. Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. *J Clin Epidemiol.* 2004; 57:1153–60. [PubMed: 15567631]
16. 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; 8:283–91. [PubMed: 15207508]
17. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis.* 2005; 64:29–33. [PubMed: 15208174]
18. Hanley MA, Jensen MP, Ehde DM, Robinson LR, Cardenas DD, Turner JA, et al. Clinically significant change in pain intensity ratings in persons with spinal cord injury or amputation. *Clin J Pain.* 2006; 22:25–31. [PubMed: 16340590]

19. van der Roer N, Ostelo RWJG, Bekkering GE, van Tulder MW, de Vet HCW. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine*. 2006; 31:578–82. [PubMed: 16508555]
20. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR, et al. Minimally important difference in diffuse systemic sclerosis: Results from the D-penicillamine study. *Ann Rheum Dis*. 2006; 65:1325–9. [PubMed: 16540546]
21. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskeletal Disorders*. 2006; 7:82. [PubMed: 17064410]
22. Kovacs FM, Abraira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine*. 2007; 32:2915–20. [PubMed: 18246018]
23. Pope JE, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the Health Assessment Questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *J Rheumatol*. 2009; 36:254–9. [PubMed: 19132791]
24. Spiegel B, Bolus R, Harris LA, Lucak S, Naliboff B, Esrailian E, et al. Measuring irritable bowel syndrome patient-reported outcomes with an abdominal pain numeric rating scale. *Aliment Pharmacol Ther*. 2009; 30:1159–70. [PubMed: 19751360]
25. Aletaha D, Funovitis J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. *Arthritis Care Res*. 2009; 61:313–20.
26. Farrar JT, Pritchett YL, Robinson M, Prakash A, Chappell A. The clinical importance of changes in the 0 to 10 numeric rating scale for worst, least, and average pain intensity: Analyses of data from clinical trials of duloxetine in pain disorders. *J Pain*. 2010; 11:109–18. [PubMed: 19665938]
27. Browne JP, van der Meulen JH, Lewsey JD, Lamping DL, Black N. Mathematical coupling may account for the association between baseline severity and minimally important difference values. *J Clin Epidemiol*. 2010; 63:865–74. [PubMed: 20172689]
28. Demoulin C, Ostelo R, Knottnerus JA, Smeets RJEM. What factors influence the measurement properties of the Roland-Morris disability questionnaire? *Eur J Pain*. 2010; 14:200–6. [PubMed: 19443246]
29. Emschhoff R, Bartram S, Emschhoff I. Clinically important difference thresholds of the visual analog scale: A conceptual model for identifying meaningful intraindividual changes for pain intensity. *Pain*. 2011; 152:2277–82. [PubMed: 21726939]
30. Adamchic I, Tass PA, Langguth B, Haputmann C, Koller M, Schecklmann M, et al. Linking the Tinnitus Questionnaire and the subjective clinical global impression: Which differences are clinically important? *Health Qual Life Outcomes*. 2012; 10:79. [PubMed: 22781703]
31. Wang Y-C, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011; 91:675–88. [PubMed: 21372203]
32. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics*. 2000; 18:419–23. [PubMed: 11151395]
33. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum*. 1988; 31:315–24. [PubMed: 3358796]
34. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989; 10:407–15. [PubMed: 2691207]
35. Ward MM, Marx AS, Barry NN. Identification of clinically important changes in health status using receiver operating characteristic curves. *J Clin Epidemiol*. 2000; 53:279–84. [PubMed: 10760638]
36. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. 1992; 30:917–25. [PubMed: 1405797]

What is new?

- This is the first study to test if patients who comprise the subgroup with a high MCII on one measure of disease activity also constitute subgroups with high MCII for other measures of disease activity.
- Comparing the concordance of patient responses across measures, MCII in the high baseline subgroups segregate with the measure rather than with particular patients.
- Variation in the MCII with the baseline value is attributable to differences in maximum possible changes and opportunities for misjudgments at different baseline values, and was similar for three measures of RA activity.
- A single MCII can be estimated for all patients provided they meet a minimum level of disease activity or severity.

Figure 1.

Measured changes in the patient global assessment by the value of the patient global assessment at baseline (0–100). Subjects were grouped into 10 subgroups by their baseline value, and plotted at the midpoint of the interval. Positive changes indicate worsening in the patient global assessment between visits; negative changes indicate improvement in the patient global assessment between visits. Plus signs represent subjects who judged themselves as improved; circles represent subjects who judged themselves as not improved.

Figure 2.

Measured changes in the swollen joint count by the value of the swollen joint count at baseline. Subjects were grouped into 7 subgroups by their baseline value. Positive changes indicate worsening in the swollen joint count between visits; negative changes indicate improvement. Plus signs represent subjects who judged themselves as improved; circles represent subjects who judged themselves as not improved.

Figure 3.

Measured changes in the walking time by quintile of the walking time at baseline, in seconds. Positive changes indicate worsening in the walking time between visits; negative changes indicate improvement. Plus signs represent subjects who judged themselves as improved; circles represent subjects who judged themselves as not improved.

Figure 4.

Distribution of swollen joint count at baseline (horizontal axis) and walking time at baseline (vertical axis). Subjects with a patient global assessment of 90–100 at baseline are denoted by dots, those with a patient global assessment of 80–89 at baseline are denoted by triangles, and those with a patient global assessment < 80 at baseline are denoted by open circles.

Figure 5.

A. Distribution of subjects who judged they had improvement in their swollen joints, by the number of swollen joints at baseline. Those with baseline values of patient global assessment of 80–100 are represented by dots, and those with baseline values of patient global assessment of 0–79 are represented by open circles. B. Distribution of subjects who judged they had improvement in functional ability, by their walking time at baseline. Those with baseline values of patient global assessment of 80–100 are represented by dots, and those with baseline values of patient global assessment of 0–79 are represented by open circles.

Table 1

Minimal clinically important improvement (MCII) estimates and standardized response means (SRM) in subgroups by the baseline value of the patient global assessment, swollen joint count, and walking time.

Patient Global Assessment				Swollen Joint Count						Walking Time			
Baseline value	N	MCII	SRM	Baseline value	N	MCII	SRM	Joint swelling as Top 3 Concern (%)	Baseline value	N	MCII	SRM	Functional Difficulty as Top 3 Concern (%)
0 – 9	7	+1	+0.27	0 – 5	21	+2	+0.13	55	0 – 8	50	0	0.22	14
10 – 19	17	+11	+0.20	6 – 10	59	–5	–0.61	59	8.1 – 9.3	32	–1.1	–0.44	24
20 – 29	24	–10	–0.03	11 – 15	52	–7	–0.71	61	9.4 – 11.9	51	–1.0	–0.41	25
30 – 39	21	–7	–0.06	16 – 20	47	–8	–0.78	68	12.0 – 14.9	48	–4.0	–0.65	23
40 – 49	19	–5	–0.54	21 – 25	39	–8	–1.03	49	15.0 or more	51	–7.0	–0.97	19
50 – 59	51	–5	–0.69	26 – 30	15	–11	–1.07	53					
60 – 69	33	–29	–0.88	31 or more	17	–22	–2.35	41					
70 – 79	29	–28	–1.54										
80 – 89	22	–38	–1.57										
90 – 100	27	–42	–1.51										