



LUND UNIVERSITY

Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement

Davis, A. M.; Perruccio, A. V.; Lohmander, Stefan

Published in:
Osteoarthritis and Cartilage

DOI:
[10.1016/j.joca.2012.02.005](https://doi.org/10.1016/j.joca.2012.02.005)

2012

[Link to publication](#)

Citation for published version (APA):

Davis, A. M., Perruccio, A. V., & Lohmander, S. (2012). Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement. *Osteoarthritis and Cartilage*, 20(5), 364-367. <https://doi.org/10.1016/j.joca.2012.02.005>

Total number of authors:
3

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

**Minimally Clinically Important Improvement:
All non-responders are not really non-responders
An illustration from total knee replacement**

Corresponding Author:

Aileen M. Davis PhD
MP-11, Room 322
Toronto Western Research Institute
399 Bathurst Street
Toronto, ON Canada M5T 2S8
Email: adavis@uhnresearch.ca
Telephone: 416-603-5543
Fax: 416-603-6288

Key words: minimal clinically important difference, responder, cohort, osteoarthritis

Brief Report

Introduction and Summary

Patient-reported outcomes (PROs) are accepted endpoints in the evaluation of patient treatment. Attention has focused on defining the minimal clinically important improvement/change (MCID) of PROs as a way to identify response to treatment. Importantly, the number needed to treat that is commonly used in similar circumstances is based on the responder rate.

Goldsmith et al introduced the concept of the MCID based on expert clinical opinion in 1993.[1] Subsequently, Jaeschke et al.[2] defined the MCID as “...the smallest difference in score in the domain of interest that patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management.” Their methods included the use of a patient global rating where the patient indicates how much better or worse they perceive they are on a scale ranging from -7, a great deal worse to +7, a great deal better. The mean score of those who rated themselves as +1, a little better, were perceived to have achieved a MCID. This has been labeled the anchor-based approach. Other authors have slightly different definitions of clinically important difference[3] and still others have used variations of the anchor-based approach.[4] The recent Federal Drug Administration guidelines for PROs affirmed that anchor-based methodology was required in reporting the proportion responding to treatment in the evaluation of all medical devices and

drugs.[5] A recent review by King details the definitions and methods for determining the MCID.[4]

It is accepted that a PRO does not have a single MCID but rather that it is an interaction of the outcome measure and the context in which it is used. As such, a given PRO has a range of MCIDs and the value varies depending on the patient group and the intervention.[4]

Additionally, as reviewed by King[4] several authors have identified methodological challenges in determining important change. These challenges include but are not limited to reported values for important change that are less than measurement error[6] and variability in the magnitude of perceived important change depending on the baseline score.[7] However, authors have reported all of those not achieving the MCID, irrespective of baseline score, as non-responders without considering whether the MCID was not achieved due to a ceiling effect. Further, the potential influence of regression to the mean phenomenon has not been considered Recent studies of cohorts of hip[8] and knee replacement[9] are examples.

Much of the developmental work and application of MCID has been in the context of clinical trials.[2-4] As such, this work has occurred in the context of pre-specified eligibility criteria, often inclusive of a threshold score for the PRO that will be used as the outcome. Recently, researchers also have defined responders and evaluated factors associated with achieving MCID in patient cohorts where such eligibility criteria have not been used.[8, 9] These works have not considered how baseline scores less than the magnitude of the MCID impact interpretation of response. We here illustrate in a group of people with total knee replacement for osteoarthritis that, in such cohort studies, there are people who respond, those who do not respond and a third group who are misclassified as non-responders because their baseline score does not allow achievement of important change.

Methods, Results and Discussion

To illustrate the impact of classifying as non-responders those who cannot achieve a MCID due to their pre-intervention PRO score, we chose a cohort of people who had primary total knee replacement (TKR) for osteoarthritis. TKR has been shown to be an effective surgical procedure, with a low complication rate, for relieving pain and improving function in people with moderate to severe knee OA when medical management has failed.[10] The Western Ontario McMaster Universities' Osteoarthritis Outcome Measure (WOMAC) is a commonly used PRO measure in this patient group. This work used the WOMAC Likert version 3.0 pain and function subscales, scored 0 to 20 and 0-68 respectively with zero representing no pain or functional difficulties respectively, as a clinically important difference value for the subscales has been published for people having TKR in the geographic region of our sample.[11] Similar confirmatory data from Spain have been published.[12]

The present work focuses on the WOMAC pain and function data from pre-surgery and 12 months post-surgery from a sample of 494 people who had TKR; 483 of these had pre-surgery and one year WOMAC pain scores and 486 had function scores. The patients were initially recruited to a longitudinal study designed to evaluate the trajectory of recovery and the inter-relationships of symptoms, activity limitations and participation restrictions in people undergoing primary TKR for osteoarthritis of the knee.[13] Eligibility criteria included age of 18 years or older who had English fluency sufficient for completion of patient-reported outcomes and consent to participate. Participants, recruited from five tertiary care centres, completed a battery of questionnaires pre-surgery and 12 months post surgery.

The sample demographics (Table 1) are typical of those receiving TKR and, on average, there were large and statistically significant improvements for the group from pre-surgery to one-year follow-up. Participants were classified as having achieved an important improvement on the pain subscale based on achieving a change score of at least 7 points (of 20 points total) for WOMAC pain from pre-surgery to one year post-surgery, a time when recovery is accepted to have stabilized and maximized for most people.[13] For function, participants achieved an important difference if the change score was 22 points or greater (of a total of 68) from pre-surgery to one year post-surgery. As our cohort included patients from similar hospitals, we used the one-year clinically important difference values published by Chesworth et al[11] (and used by Alzahrani et al. in their study of TKR).[9]

We calculated the proportion of people who could not achieve important change based on both the pre-surgery pain and function score (i.e., for pain with a score range of 0-20, the MCID value used was 7 and from function with a score range of 0 to 68 the MCID value used was 22 points). To demonstrate if baseline score affected achieving an important change, we calculated and compared the proportion achieving an important change in each of pain and function based on tertiles of the pre-surgery pain and function subscale score. Additionally, we evaluated whether satisfaction with outcome varied in these three groups for each of pain and function based on the response to a question asking how satisfied the individual was with the outcome of their surgery (response ranging from very dissatisfied to very satisfied). We anticipated that if people did not achieve an important improvement, considering pain and function subscales separately, there would be a tendency for them to be less satisfied with their surgical outcome assuming satisfaction is also a surrogate of efficacy. We did not evaluate important deterioration as few people having TKR (5% or less) report worsening of pain and function.[11, 12]

Response to Treatment with TKR

Of the 483 people, 54.7% (n=264) achieved an important improvement in pain (7 or more points improvement of 20) and 40.4% (195 of 486) achieved important improvement in function (22 or more points improvement of 68) at one year after surgery. Of those who did not achieve an important improvement in pain, 139 had a baseline pain score below 7 points and 155 had a baseline function score below 22 such that these individuals did not have potential to achieve important change or to be classified as a responder prior to their surgery. That is, 28.8 % (139/483) of the total sample were pre-determined prior to surgery to be a non-responder based on their pre-surgery pain score. This proportion was 31.9% for WOMAC function.

Dividing the sample into tertiles based on pre-surgery pain score, Table 2 shows that the proportion of responders increased as the baseline score increased (i.e., those with more pain and functional disability pre-surgery were more likely to achieve important improvements in pain and function). From lowest to highest pre-surgery pain tertile (high is worse pain), 30.5, 66.7 and 74.3% respectively achieved important change ($p < 0.0001$). We also note that there is a significantly higher proportion of females with worse pain at baseline ($p = 0.0127$) and that while there is a statistically significant difference in age across the tertiles, this is likely not meaningful clinically. There was no statistically significant difference in satisfaction across tertiles of baseline score. Among those unable to achieve a MCID for pain, 64.9% reported being very satisfied. The equivalent estimate for physical function was 65.2%. The proportions reported across tertiles for both outcomes were similar.

Similarly, there was a statistically significant difference ($p < 0.0001$) in the proportion who achieved an important improvement in function, 4.7, 47.5 and 67.4% respectively, with those

with more functional problems pre-surgery more likely to achieve important improvement. A higher proportion of females had poorer function pre-surgery but age was similar across tertiles. There was no statistically significant difference in satisfaction with outcome based on pre-surgery function score.

Implications and Recommendations

These data demonstrate that defining individuals as non-responders to treatment based on a single value MCID without consideration of the potential effect of the pre-intervention PRO score may result in a significant proportion of individuals being misclassified as having received no benefit from treatment. They reinforce the findings of Stratford et al. in people treated with physiotherapy for low back pain[7] and Tubach et al. in people treated medically with osteoarthritis.[14] Tubach has suggested that the MCID be calculated for the tertiles of the baseline score as a more accurate depiction of response across the range of a given PRO.[14] The broader implications of defining response to treatment using a single value MCID without consideration of baseline score, including ceiling effects and possible regression to the mean, in terms of misclassification bias will only be known as clinicians and scientists evaluate this phenomenon in additional work.

As an alternative to the MCID, The Patient Acceptable Symptom State (PASS), threshold values that must be achieved to have an acceptable outcome, is a measure of response that is less sensitive to the baseline PRO score.[15] However, the baseline score impact is predicated on whether those who have PRO scores that do not exceed the threshold are included in the sample under study (i.e. the state may be acceptable but if there is an opportunity to be even better an individual may choose additional treatment.) More recent work by Beaton et. al. evaluating response to treatment using a diagnostic testing approach demonstrated that using a combined

MCID and final status (PASS) approach had greatest accuracy (sensitivity and specificity) and clinical sensibility as compared to the MCID or PASS alone.[16] Additionally, our group showed that the relationship of prior status and change in status of one construct (e.g. pain) impacted another construct (e.g. function and or higher demand activities such as those required for social role participation)[13] such that multiple important constructs of outcome need to be considered simultaneously.

In a clinical context with individual patients, there might be a temptation to suggest that people with minimal symptoms and or disability should not be the group targeted for intervention be it joint replacement surgery or another disease and intervention. However, clinicians often are balancing a number of clinical and patient characteristics in providing advice for and implementation of treatment and PRO scores are not routinely part of the decision-making process. This is demonstrated in joint replacement where PROs including such constructs as pain and function alone have not been discriminating in deciding who is recommended to receive joint replacement; it is accepted that the decision to recommend such surgery includes additional patient considerations.[17] Therefore, clinical cohorts likely will continue to include patients with minimal symptoms and disability. Irrespective, by considering surgery for those who have only more pain and functional limitations on a measure or restricting cohorts to those who are generally worse, then by design alone a larger proportion would achieve an MCID, suggesting the intervention was favourable, but with little, if any, effect on level of patient satisfaction.

In summary, this work has demonstrated that using a single value MCID for a PRO summary score or PRO subscale to determine response to treatment when the sample includes individuals who could not achieve important improvement based on their pre-treatment status

leads to misclassification bias given that the proportion of individuals who are satisfied with their results. Future research needs to focus on advancing methodology for accurately defining criteria for response to treatment. In the interim, given that many PROs have MCID values that have been determined without considering the baseline score, we recommend creating a third category of response to identify those who cannot achieve response prior to intervention to understand how the baseline score impacts the interpretation of the study results.

Authors:

1. Aileen M. Davis PhD, Senior Scientist, Division of Health Care and Outcomes Research and Arthritis Community Research and Evaluation Unit, Toronto Western Research Institute; Professor, Departments of Physical Therapy, Rehabilitation Science, Institute of Health Policy, Management and Evaluation and Institute of Medical Science, University of Toronto, Toronto, Canada

2. Anthony V. Perruccio PhD, Scientific Research Associate, Department of Orthopaedic Surgery, Toronto Western Hospital and Assistant Professor, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

3. L. Stefan Lohmander MD PhD, Professor, Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden and the Institute of Sports Science and Clinical Biomechanics, and Department of Orthopaedics and Traumatology, University of Southern Denmark, Denmark

Contributions

Davis: study conception, analysis, drafting of manuscript, principle investigator of the original study from which the data were obtained.

Perruccio: study conception, analysis, review and approval of the final manuscript

Lohmander: study conception, review and approval of the final manuscript

Competing interests

None of the authors have any competing interests in relation to this manuscript.

Role of the Funding Source

No external funding was received in connection with this manuscript. The original study from which the secondary data analysis was conducted was funded by an unrestricted operating grant from the Canadian Institutes of Health Research (CIHR) 77518 to Davis.

Table 1 Sample characteristics (n=494) and WOMAC pain and function scores

Characteristic	
% female	65.4
Age (mean, range in years)	64.9 (35-88)
Pre-surgery WOMAC pain (mean, range)*	10.4 (0-20)
Pre-surgery WOMAC function (mean, range)*	33.8 (3-68)
One-year post surgery WOMAC pain (mean, range)	3.4 (0-18)
One-year post surgery WOMAC function (mean, range)	14.0 (0-68)

*WOMAC pain subscale possible score range is 0-20 where 0 represents no pain and function subscale possible range is 0-68 where 0 represents no functional limitations.

Table 2: Sample characteristics and achievement of clinically important improvement by tertiles of pre-surgery WOMAC pain and function score

WOMAC pain				
	Less pre-surgery pain		More pre-surgery pain	
tertile score range	Tertile 1 (0-9)	Tertile 2 (10-12)	Tertile 3 (13-20)	p-value*
n (%)	188 (38.9)	161 (33.3)	134 (27.4)	
pre-surgery pain (mean, mean percent change of total score)	6.9 (34.5%)	11.0 (55.0%)	14.6 (73.0%)	<0.0001
% female	57.6	66.5	73.1	0.0127
age (mean, range in years)	66.7 (41-88)	65.7 (38-88)	61.3 (44-87)	<0.0001
% attaining important change ($\geq 7/20$)	30.5	66.7	74.3	<0.0001
% very satisfied	64.3	62.9	66.7	0.8446
% very dissatisfied	4.1	3.5	3.6	
WOMAC function				
	Less pre-surgery functional problems		More pre-surgery functional problems	
tertile score range	Tertile 1 (3-28)	Tertile 2 (29-39)	Tertile 3 (40-68)	p-value*
n (%)	161 (33.1)	157 (32.3)	168 (34.6)	
pre-surgery function (mean, mean percent change of total score)	19.6 (28.8%)	34.1 (50.1%)	47.3 (69.5%)	<0.0001
% female	53.4	63.1	77.4	0.1375
age (mean, range in years)	66.0 (35-86)	64.8 (41-88)	63.8 (37-87)	<0.0001
% attaining important change ($\geq 22/68$)	4.7	47.5	67.4	<0.0001
% very satisfied	69.6	65.4	58.3	0.1121
% very dissatisfied	3.4	2.2	5.6	

*chi-square test, ANOVA, or Kruskal-Wallis test as appropriate to data

References

1. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 1993; 20: 561-565.
2. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407-415.
3. 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.
4. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011; 11: 171-184.
5. U.S. Department of Health and Human Services Food and Drug Administration, Center of Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). December 2009
Clinical/Medical. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. In:
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
6. 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.
7. 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.

8. Judge A, Cooper C, Williams S, Dreinhofer K, Dieppe P. Patient-reported outcomes one year after primary hip replacement in a European Collaborative Cohort. *Arthritis Care Res (Hoboken)* 2010; 62: 480-488.
9. Alzahrani K, Gandhi R, Debeer J, Petruccelli D, Mahomed N. Prevalence of clinically significant improvement following total knee replacement. *J Rheumatol* 2010; 38: 753-759.
10. Gunther KP. Surgical approaches for osteoarthritis. *Best Pract Res Clin Rheumatol* 2001; 15: 627-643.
11. Chesworth BM, Mahomed NN, Bourne RB, Davis AM. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol* 2008; 61: 907-918.
12. Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007; 15: 273-280.
13. Davis AM, Perruccio AV, Ibrahim S, Hogg-Johnson S, Wong R, Streiner DL, et al. The trajectory of recovery and the inter-relationships of symptoms, activity and participation in the first year following total hip and knee replacement. *Osteoarthritis Cartilage* 2011; 19: 1413-1421.
14. 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.
15. Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. *Arthritis Rheum* 2006; 55: 526-530.

16. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, et al. 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.
17. Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? Int J Rheum Dis 2011; 14: 175-180.